

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> :		A2	(11) International Publication Number: WO 00/61793 (43) International Publication Date: 19 October 2000 (19.10.00)
C12Q 1/18			
(21) International Application Number: PCT/EP00/03135		(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 7 April 2000 (07.04.00)			
(30) Priority Data: 99107031.9 9 April 1999 (09.04.99) EP 00102111.2 4 February 2000 (04.02.00) EP			
(71) Applicant (for all designated States except US): GPC BIOTECH AG [DE/DE]; Fraunhoferstrasse 20, D-82152 Martinsried (DE).			
(72) Inventors; and		Published	
(75) Inventors/Applicants (for US only): LOFERER, Hannes [AT/DE]; Alpenstrasse 74, D-82538 Geretsried (DE). JACOBI, Alexander [DE/DE]; Haidelweg 22, D-81241 München (DE).		Without international search report and to be republished upon receipt of that report.	
(74) Agent: VOSSIUS & PARTNER; Siebertstrasse 4, D-81675 München (DE).			

(54) Title: NOVEL METHOD FOR IDENTIFYING ANTIBACTERIAL COMPOUNDS

(57) Abstract

The present invention relates to a method for identifying an antagonist or inhibitor of the expression of a gene encoding a polypeptide essential for bacterial growth or survival as well as for an antagonist or inhibitor of said polypeptide. The invention further relates to a method for improved antagonists or inhibitors. The invention also provides an antagonist or inhibitor of the activity of said polypeptide. The invention is further related to a method for producing a therapeutic agent in a composition comprising said antagonist or inhibitor. Furthermore, the invention is related to the use of the polypeptide and the antagonist or inhibitor as well as to a method to identify a surrogate marker.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

### Novel method for identifying antibacterial compounds

The present invention relates to a method for identifying an antagonist or inhibitor of the expression of a gene encoding a polypeptide essential for bacterial growth or survival as well as for an antagonist or inhibitor of said polypeptide. The invention further relates to a method for improved antagonists or inhibitors. The invention also provides an antagonist or inhibitor of the activity of said polypeptide. The invention is further related to a method for producing a composition comprising said antagonist or inhibitor. Furthermore, the invention is related to the use of the polypeptide and the antagonist or inhibitor as well as to a method to identify a surrogate marker.

Several documents are cited throughout the text of this specification. Each of the documents cited herein (including any manufacturer's specifications, instructions, etc.) are hereby incorporated by reference; however, there is no admission that any document cited is indeed prior art of the present invention.

Since the beginning of the 1980s, a new trend has been observed in the industrialized countries. On the one hand, resistances to antibiotics have increased, which make it difficult or even impossible to treat many of the disease-causing agents. On the other hand, new infectious diseases, which had been unknown up to now, arise, and old diseases return. For example, diphtheria and tuberculosis are old epidemics and increasingly surmounting in many different parts of the world. Especially tuberculosis (TB), a chronic infectious disease that is generally caused by infection with *Mycobacterium tuberculosis*, is a disease of major concern. Each year, 8 to 10 million new cases of TB are described, and, causing more than three million deaths per year, TB is a major disease in developing countries as well as an increasing problem in developed areas of the world due to, for example, antibiotic resistance.

Additionally, *M. bovis* BCG vaccination has failed to protect against TB in several trials (WHO, Tech. Rep. Ser. (1980), 651, 1-15) for reasons that are not entirely clear (Fine, *Tubercle* 65 (1984), 137-153). It has been shown that the vaccine strain of *M. bovis* BCG only confers protection against the severe form of miliary tuberculosis in children (Fine, *Lancet* 346 (1995), 1339-1345). In contrast, its protective capacity against the most common form, pulmonary tuberculosis in adults, is low and highly variable (Colditz (1994), *JAMA* 271, 698).

The causes for this new trend are complex: mainly, the increasing number of antibiotic applications in medicine and agriculture often combined with an improper and uncontrolled use, helps to establish resistant organisms and generate the threat of bacterial infections resistant to all available therapies.

Conventional techniques of developing antibiotics, i.e. synthesis of candidate substances and screening for antibacterial substances, even though speeded up by several orders of magnitude by the use of combinatorial approaches in recent years (e.g. US5324483, US5545568), are still too inefficient as they involve multiple screening steps of hundreds or thousands of more or less randomly chosen substances for efficiency in combating various infectious agents.

Therefore, it is a major concern to fight the growing number of bacterial infections due to an increased frequency of multiple antibiotic resistances and to improve the available antibacterial therapies.

Thus, the technical problem underlying the present invention was to provide a method and means for the development of an additional effective antibacterial therapy of infected humans and animals that can be used for the treatment of a broad spectrum of bacterial infections or diseases or disorders related to bacterial infections. The solution to this technical problem is achieved by providing the embodiments characterized in the claims.

Accordingly, the present invention relates to a method for identifying an antagonist or inhibitor of the expression of a gene encoding a polypeptide essential for bacterial growth wherein said gene is selected from the group consisting of *ygbB*, *yfhC*, *yacE*, *ychB*, *yejD*, *yrfl*, *yggJ*, *yjeE*, *yiaO*, *prdC*, *yhbC*, *ygbP*, *ybeY*, *gcpE*, *kdtB*, *pfs*, *ycaJ*, *b1808*, *yeaA*, *yagF*, *b1983*, *yidD*, *yceG* and/or *yjbC* the sequence

of said genes being shown in Fig. 1, or a fragment or derivative or ortholog thereof, said method comprising the steps of

- (a) testing a candidate antagonist or inhibitor or a sample comprising a plurality of said candidate antagonists or inhibitors for the inhibition or reduction of transcription of said gene or a fragment or derivative thereof; or
- (b) testing a candidate antagonist or inhibitor or a sample comprising a plurality of said candidate antagonists or inhibitors for the inhibition or reduction of translation of mRNA transcribed from said gene or a fragment or derivative thereof; and
- (c) identifying an antagonist or inhibitor or a sample comprising a plurality of said candidate antagonists or inhibitors that tests positive in step (a) and/or (b).

The term "antagonist" or "inhibitor" as used herein means naturally occurring and synthetic compounds capable of counteracting or inhibiting an activity of a gene or gene product or interactions of the gene or gene product with other genes or gene products. Determining whether a compound is capable of inhibiting or counteracting specific gene expression can be done, for example, by Northern blot analysis, Western blot analysis or proteome analysis. It can further be done by monitoring the phenotypic characteristics of a bacterial cell contacted with the compounds and compare it to that of a wild-type cell. In an additional embodiment, said characteristics may be compared to that of a cell contacted with a compound which is either known to be capable or incapable of suppressing or activating the protein or gene, respectively, according to the invention. For example, the bacterial cell can be a transgenic cell and the phenotypic characteristics comprises a readout system. Further examples of determining whether a compound is capable of inhibiting or counteracting specific gene expression are described below.

The term "expression" means the production of a protein or nucleotide sequence in a cell. However, said term also includes expression of the protein in a cell-free system. It includes transcription into an RNA product, and/or translation into a polypeptide from a DNA encoding that product.

The term "transcription" as used herein means a DNA template dependent synthesis of a ribonucleic acid polymer encoding a polypeptide or a regulatory

sequence. The term "translation" as used herein means the polymerization of a polypeptide that is encoded by an RNA molecule by a protein complex.

As used in accordance with the present invention, the term "fragment or derivative" denotes any variant the amino acid or nucleotide sequence of which deviates in its primary structure, e.g., in sequence composition or in length as well as to analogue components. For example, one or more amino acids of a polypeptide may be replaced in said fragment or derivative as long as the modified polypeptides remain functionally equivalent to their described counterparts. The term "fragment or derivative" further denotes compounds analog to an antagonist or inhibitor that should have a stabilized electronic configuration and molecular conformation that allows key functional groups to be presented to the mentioned polypeptide in substantially the same way as the antagonist and inhibitor. The variant of the polypeptide may be a naturally occurring allelic variant of the polypeptide or non-naturally occurring variants of those polynucleotides.

The term "orthologs" as used herein means homologous sequences in different species that evolved from a common ancestral gene by speciation. Normally, orthologs retain the same function in the course of evolution. However, orthologous genes may or may not be responsible for a similar function (see, e.g., the glossary of the "Trends Guide to Bioinformatics", Trends Supplement 1998, Elsevier Science). Orthologous genes, nucleic acids or proteins comprise genes, nucleic acids or proteins which have one or more sequences or structural motifs in common. For example, the sequence motifs of proteins can comprise short, i.e. repetitive sequences or amino acid positions conserved in the primary structure and/or conserved in higher protein structures, e.g. secondary or tertiary structure. Orthologous nucleic acids or genes can comprise molecules having short stretches of one or more homologous (same or similar) sequences, for example protein binding boxes or structure forming boxes. Methods for the identification of a candidate ortholog of a gene or polypeptide described herein are known to those skilled in the art and are described for example in Sambrook et al. (1989), Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York, or Ausubel (1994), Current Protocols in Mol. Biol.. The person skilled in the art knows how to identify orthologous genes, nucleic acids or polypeptides by

computer supported analysis (e.g. BLAST) of known sequences and its interpretation.

The terms "gene", "polynucleotide", "nucleic acid sequence", "nucleotide sequence", "DNA sequence" or "nucleic acid molecule" as used herein refer to polymeric forms of nucleotides of any length, either ribonucleotides or deoxyribonucleotides and only to the primary structure of the molecule. Thus, these terms include double- and single-stranded DNA, and RNA. They also include known types of modifications, for example, methylation, "caps" substitution of one or more of the naturally occurring nucleotides with an analog. Preferably, the DNA sequence of the invention comprises a coding sequence encoding at least the mature form of the above defined protein, i.e. the protein which is posttranslationally processed in its biologically active form, for example due to cleavage of leader or secretory sequences or a proprotein sequence or other natural proteolytic cleavage points.

The term "plurality of candidate antagonists or inhibitors" is to be understood as a plurality of substances which may or may not be identical.

Said antagonists or inhibitors or plurality of candidate antagonists or inhibitors may be chemically synthesized or microbiologically produced and/or comprised in, for example, samples, e.g., cell extracts from, e.g., plants, animals or microorganisms. Furthermore, said compound(s) may be known in the art but hitherto not known to be capable of suppressing or inhibiting said polypeptide. The reaction mixture may be a cell free extract or may comprise a cell or tissue culture. Suitable set ups for the method of the invention are known to the person skilled in the art and are, for example, generally described in Alberts et al., Molecular Biology of the Cell, third edition (1994), in particular Chapter 17. The plurality of compounds may be, e.g., added to the reaction mixture, culture medium, injected into the cell or sprayed onto the plant.

By combining computational processing of genomic information with microbial genetics, the inventors have been able to identify 24 *E. coli* essential genes and their respective orthologs (Fig. 3) that fulfill several criteria for being attractive antibacterial targets: hypothetical open reading frames, coding for essential functions (mutation is lethal for growth in rich media), broad conservation (orthologs are present in a wide range of bacteria including *H. influenza*, *S.*

*pneumoniae*, *H. pylori*, and *B. burgdorferi*) (Fig. 3) and low toxicity potential in higher organisms (mostly no orthologs are identified in the simple eukaryote *S. cerevisiae*). Thus, an antagonist or inhibitor of the expression of such an essential gene or of its function provides the key for an antibacterial therapy. The inventors assume that said antagonist or inhibitor stops or reduces bacterial growth and/or mediates bacterial death.

Thus, the method of the present invention provides the options of development of new broad spectrum antibiotics against new pharmaceutical important targets. The findings of the present invention are particularly important in view of the drawbacks of the present forms of treatment of bacterial infections, diseases and disorders related to bacterial infections.

In line with the above, the present invention also relates to a method for testing a candidate antagonist or inhibitor of a polypeptide or mRNA essential for bacterial growth or survival encoded by a gene selected from the group consisting of *ygbB*, *yfhC*, *yacE*, *ychB*, *yejD*, *yrfl*, *yggi*, *yjeE*, *yiaO*, *prdC*, *yhbC*, *ygbP*, *ybeY*, *gcpE*, *kdtB*, *pfs*, *ycaJ*, *b1808*, *yeaA*, *yagF*, *b1983*, *yidD*, *yceG* and/or *yjbC* or a fragment, derivative or ortholog thereof comprising the steps of

- (a) contacting a bacterial cell with candidate antagonist or inhibitor or a sample comprising a plurality of said candidate antagonists or inhibitors; and
- (b) testing whether said contacting leads to cell growth inhibition and/or cell death.

In a further embodiment, the present invention relates to a method for testing a candidate antagonist or inhibitor of the function of a gene essential for bacterial growth or survival wherein said gene is selected from the group consisting of *ygbB*, *yfhC*, *yacE*, *ychB*, *yejD*, *yrfl*, *yggi*, *yjeE*, *yiaO*, *prdC*, *yhbC*, *ygbP*, *ybeY*, *gcpE*, *kdtB*, *pfs*, *ycaJ*, *b1808*, *yeaA*, *yagF*, *b1983*, *yidD*, *yceG* and/or *yjbC* or a fragment, derivative or ortholog thereof, comprising the steps of

- (a) contacting a bacterial cell comprising said gene with a candidate antagonist or inhibitor or a sample comprising a plurality of said candidate antagonists or inhibitors; and

(b) testing whether said contacting leads to cell growth inhibition and/or cell death.

Bacteria, for which was shown that a gene as mentioned above expressed is essential, can be used in a proliferation assay to identify both ligands and potential antagonists or inhibitors to said polypeptide encoded by said essential gene. For example, *E. coli* are grown in culture medium and incorporation of DNA precursors such as  $^3\text{H}$ -thymidine or 5-bromo-2'-deoxyuridine (BrdU) is monitored as a parameter for DNA synthesis and cellular proliferation. Cells which have incorporated BrdU into DNA can be detected using a monoclonal antibody against BrdU and measured by an enzyme or fluorochrome-conjugated second antibody. The reaction is quantitated by fluorimetry or by spectrophotometry. The ability of the compound to be screened to inhibit proliferation may then be quantified. Further methods to determine growth and proliferation of bacteria are well known in the art, for example in Drews, Mikrobiol. Praktikum, Berlin, 1976.

Preferably, the antagonist or inhibitor binds to the gene product, i.e. the RNA or polypeptide, specifically encoded by said gene.

For example, a candidate antagonist or inhibitor not known to be capable of binding to an polypeptide encoded by a essential gene as described above can be tested to bind thereto comprising contacting a bacterial cell comprising an isolated molecule encoding said polypeptide with a candidate antagonist or inhibitor under conditions permitting binding of ligands known to bind thereto, detecting the presence of any bound ligand, and thereby determining whether such candidate antagonist or inhibitor inhibits the binding of a ligand to a polypeptide as described above.

Proteins that bind to a polypeptide as described above and might inhibit or counteract to said polypeptide can be "captured" using the yeast two-hybrid system (Fields, Nature 340 (1989), 245-246). A modified version of the yeast two-hybrid system has been described by Roger Brent and his colleagues (Gyuris, Cell 75 (1993), 791-803; Zervos, Cell 72 (1993), 223-232). Briefly, a domain of the polypeptide is used as bait for binding compounds. Positives are then selected by their ability to grow on plates lacking leucine, and then further tested for their ability to turn blue on plates with X-gal, as previously described in great detail (Gyuris, supra; WO 95/31544). Once amino acid sequences are identified which

bind to a polypeptide essential for bacterial growth or survival, these sequences can be screened for antagonist activity using, for example, the proliferation assay described above or used for screening for antagonists of said binding.

Another assay which can be performed to identify inhibitors and antagonists involves the use of combinatorial chemistry to produce random peptides which then can be screened for both binding affinity and antagonist effects. One such assay has recently been performed using random peptides expressed on the surface of a bacteriophage (Wu (1996), Nature Biotechnology 14, 429-431).

In a preferred embodiment of the method of the present invention said method further comprises identifying an antagonist or inhibitor optionally from said sample of candidate antagonists or inhibitors.

If a sample contains a candidate antagonist or inhibitor, or a plurality of candidate antagonists or inhibitors, as identified in the method of the invention, then it is either possible to isolate the candidate antagonists or inhibitors from the original sample identified as containing the compound capable of suppressing or inhibiting bacterial growth or survival, or one can further subdivide the original sample, for example, if it consists of a plurality of different candidate antagonists or inhibitors, so as to reduce the number of different substances per sample and repeat the method with the subdivisions of the original sample. Depending on the complexity of the samples, the steps described above can be performed several times, preferably until the sample identified according to the method of the invention only comprises a limited number of or only one substance(s). Preferably said sample comprises substances of similar chemical and/or physical properties, and most preferably said substances are identical. As regards the identification of candidate antagonists or inhibitors by any of the above-referenced embodiments of the invention, a variety of formats or tools is available to the person skilled in the art. Thus, several methods are known to the person skilled in the art for producing and screening large libraries to identify compounds having specific affinity for a target. These methods include the phage-display method in which randomized peptides are displayed from phage and screened by affinity chromatography to an immobilized receptor; see, e.g., WO 91/17271, WO 92/01047, US-A-5,223,409. In another approach, combinatorial libraries of polymers immobilized on a chip are synthesized using photolithography; see, e.g., US-A-5,143,854, WO 90/15070

and WO 92/10092. The immobilized polymers are contacted with a labeled receptor and scanned for label to identify polymers binding to the receptor. The synthesis and screening of peptide libraries on continuous cellulose membrane supports that can be used for identifying binding ligands of the polypeptide of the invention and thus possible inhibitors and antagonists is described, for example, in Kramer, Methods Mol. Biol. 87 (1998), 25-39. This method can also be used, for example, for determining the binding sites and the recognition motifs in the polypeptide as described above. In like manner, the substrate specificity of the DnaK chaperon was determined and the contact sites between human interleukin-6 and its receptor; see Rüdiger, EMBO J. 16 (1997), 1501-1507 and Weiergraber, FEBS Lett. 379 (1996), 122-126, respectively. Furthermore, the above-mentioned methods can be used for the construction of binding supertopes derived from the polypeptide of the invention. A similar approach was successfully described for peptide antigens of the anti-p24 (HIV-1) monoclonal antibody; see Kramer, Cell 91 (1997), 799-809. A general route to fingerprint analyses of peptide-antibody interactions using the clustered amino acid peptide library was described in Kramer, Mol. Immunol. 32 (1995), 459-465. In addition, antagonists or inhibitors of a polypeptide described above can be derived and identified from monoclonal antibodies that specifically react with said polypeptide in accordance with the methods as described in Doring, Mol. Immunol. 31 (1994), 1059-1067.

More recently, WO 98/25146 described further methods for screening libraries of complexes for compounds having a desired property, especially, the capacity to agonize, bind to, or antagonize a polypeptide or its cellular receptor. The complexes in such libraries comprise a compound under test, a tag recording at least one step in synthesis of the compound, and a tether susceptible to modification by a reporter molecule. Modification of the tether is used to signify that a complex contains a compound having a desired property. The tag can be decoded to reveal at least one step in the synthesis of such a compound. Other methods for identifying compounds which interact with the proteins according to the invention or nucleic acid molecules encoding such molecules are, for example, the in vitro screening with the phage display system as well as filter binding assays or "real time" measuring of interaction using, for example, the BIACore apparatus (Pharmacia).

All these methods can be used in accordance with the present invention to identify antagonists and inhibitors of the polypeptide of the invention.

Additionally, the present invention relates in a preferred embodiment to a method comprising improving inhibitors or antagonists identified by peptidomimetics or by applying phage display or combinatorial library technique step(s). Peptidomimetics, phage display and combinatorial library techniques are well-known in the art and can be applied by the person skilled in the art without further ado to the improvement of the antagonist or inhibitor that is identified by the basic method referred to herein above.

Methods for the generation and use of peptidomimetic combinatorial libraries are described in the prior art, for example in Ostresh, Methods In Enzymology 267 (1996), 220-236; Dosner, Bioorg. Med. Chem. 4 (1996), 709-715; Beeley, Trends Biotechn. 12 (1994), 213-216; al-Obeidi, Mol. Biotechn. 9 (1998), 205-223; Wiley, Med. Res. Rev. 13 (1993), 327-384; Bohm, J. Comput. Aided Mol. Des. 10 (1996), 265-272; and Hruby, Biopolymers 43 (1997), 219-266.

Various sources for the basic structure of such an antagonist or inhibitor can be employed and comprise, for example, mimetic analogs of the polypeptide of the invention. Mimetic analogs of the polypeptide of the invention or biologically active fragments thereof can be generated by, for example, substituting the amino acids that are expected to be essential for the biological activity with, e.g., stereoisomers, i.e. D-amino acids; see e.g., Tsukida, J. Med. Chem. 40 (1997), 3534-3541. Furthermore, in case fragments are used for the design of biologically active analogs pro-mimetic components can be incorporated into a peptide to reestablish at least some of the conformational properties that may have been lost upon removal of part of the original polypeptide; see, e.g., Nachman, Regul. Pept. 57 (1995), 359-370. Furthermore, the polypeptide can be used to identify synthetic chemical peptide mimetics that bind to or can function as a ligand, substrate, binding partner or the receptor of the polypeptide as effectively as does the natural polypeptide; see, e.g., Engleman, J. Clin. Invest. 99 (1997), 2284-2292.

The structure-based design and synthesis of low-molecular-weight synthetic molecules that mimic the activity of the native biological polypeptide is further

described in, e.g., Dowd, *Nature Biotechnol.* 16 (1998), 190-195; Kieber-Emmons, *Current Opinion Biotechnol.* 8 (1997), 435-441; Moore, *Proc. West Pharmacol. Soc.* 40 (1997), 115-119; Mathews, *Proc. West Pharmacol. Soc.* 40 (1997), 121-125; Mukhija, *European J. Biochem.* 254 (1998), 433-438.

It is also well known to the person skilled in the art, that it is possible to design, synthesize and evaluate mimetics of small organic compounds that, for example, can act as a substrate or ligand to a polypeptide as encoded by the essential gene as identified above. For example, it has been described that D-glucose mimetics of hapalosin exhibited similar efficiency as hapalosin in antagonizing multidrug resistance assistance-associated protein in cytotoxicity; see Dinh, *J. Med. Chem.* 41 (1998), 981-987.

The essential gene described above or the RNA encoded thereof, as has been described above, can also serve as a target for antagonists or inhibitors. Antagonists may comprise, for example, proteins that bind to the mRNA of said gene, thereby destabilizing the native conformation of the mRNA and disturbing transcription and/or translation. Furthermore, methods are described in the literature for identifying nucleic acid molecules such as an RNA fragment that mimics the structure of a defined or undefined target RNA molecule to which a compound binds inside of a cell resulting in retardation of cell growth or cell death; see, e.g., WO 98/18947 and references cited therein. These nucleic acid molecules can be used for identifying unknown compounds of pharmaceutical and/or agricultural interest, and for identifying unknown RNA targets for use in treating a disease. These methods and compositions can be used in screening for novel antibiotics, bacteriostatics, or modifications thereof or for identifying compounds useful to alter expression levels of proteins encoded by a nucleic acid molecule. Alternatively, for example, the conformational structure of the RNA fragment which mimics the binding site can be employed in rational drug design to modify known antibiotics to make them bind more avidly to the target. One such methodology is nuclear magnetic resonance (NMR), which is useful to identify drug and RNA conformational structures. Still other methods are, for example, the drug design methods as described in WO 95/35367, US-A-5,322,933, where the crystal structure of the RNA fragment can be deduced and computer programs are utilized to design novel binding compounds which can act as antibiotics.

The candidate antagonists and inhibitors which can be tested and identified according to a method of the invention may be taken from expression libraries, e.g., cDNA expression libraries, peptides, proteins, nucleic acids, antibodies, small organic compounds, hormones, peptidomimetics, PNAs or the like (Milner, *Nature Medicine* 1 (1995), 879-880; Hupp, *Cell* 83 (1995), 237-245; Gibbs, *Cell* 79 (1994), 193-198 and references cited *supra*). Furthermore, genes encoding a putative regulator of an essential bacterial protein and/or which exert their effects up- or downstream said protein may be identified using, for example, insertion mutagenesis using, for example, gene targeting vectors known in the art (see, e.g., Hayashi, *Science* 258 (1992), 1350-1353; Fritze and Walden, *Gene activation by T-DNA tagging*. In *Methods in Molecular biology* 44 (Gartland, K.M.A. and Davey, M.R., eds). Totowa: Human Press (1995), 281-294) or transposon tagging (Chandlee, *Physiologia Plantarum* 78 (1990), 105-115). Said compounds can also be functional derivatives or analogues of known inhibitors or antagonists. Such useful compounds can be for example transacting factors which bind an above-described polypeptide. Identification of transacting factors can be carried out using standard methods in the art (see, e.g., Sambrook, *supra*, and Ausubel, *supra*). To determine whether a protein binds to the protein or regulatory sequence of the invention, standard native gel-shift analyses can be carried out. In order to identify a transacting factor which binds to the protein or regulatory sequence of the invention, the protein or regulatory sequence of the invention can be used as an affinity reagent in standard protein purification methods, or as a probe for screening an expression library. The identification of nucleic acid molecules which encode proteins which interact with the polypeptide described above can also be achieved, for example, as described in Scofield (*Science* 274 (1996), 2063-2065) by use of the so-called yeast "two-hybrid system"; see also the appended example. In this system, e.g., the protein encoded by the nucleic acid molecules identified in this invention or a smaller part thereof is linked to the DNA-binding domain of the GAL4 transcription factor. A yeast strain expressing this fusion gene and comprising a lacZ reporter gene driven by an appropriate promoter, which is recognized by the GAL4 or LexA transcription factor, is transformed with a library of cDNAs which will express plant genes or fragments thereof fused to an activation domain. Thus, if a peptide encoded by one of the cDNAs is able to interact with the fusion peptide comprising a peptide of a protein of

the invention, the complex is able to direct expression of the reporter gene. In this way the nucleic acid molecules and the encoded peptide can be used to identify peptides and proteins interacting with the polypeptide described above. It is apparent to the person skilled in the art that this and similar systems may then further be exploited for the identification of inhibitors or antagonists of the polypeptide.

Once the transacting factor is identified, modulation of its binding to or regulation of expression of the polypeptide described above can be pursued, beginning with, for example, screening for inhibitors against the binding of the transacting factor to the protein specified in accordance with the present invention. Inhibition of bacterial growth could then be achieved by applying the transacting factor (or its inhibitor). In addition, if the active form of the transacting factor is a dimer, dominant-negative mutants of the transacting factor could be made in order to inhibit its activity.

Thus, the present invention also relates to the use of the polypeptide as defined above for the identification of antagonists or inhibitors of a polypeptide essential for bacterial growth or survival.

In another embodiment, the present invention relates to a method for designing an improved antagonist or inhibitor for the treatment of a bacterial infection or disorder or disease related to a bacterial infection comprising the steps of

- (a) identification of the binding site of an antagonist or inhibitor to the polypeptide ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC, the sequence of said genes being shown in Fig. 1, or identified according to the method of the present invention, by site-directed mutagenesis and chimeric polypeptide studies;
- (b) molecular modeling of both the binding site of said antagonist or inhibitor and the structure of said polypeptide; and
- (c) modification of said antagonist or inhibitor to improve its binding specificity or affinity for the polypeptide.

Biological assays as described above or other assays such as assays based on crystallography or NMR may be employed to assess the specificity or potency of

the antagonist or inhibitor wherein the decrease of one or more activities of the polypeptide may be used to monitor said specificity or potency. All techniques employed in the various steps of the method of the invention are conventional or can be derived by the person skilled in the art from conventional techniques without further ado.

For example, identification of the binding site of said antagonist or inhibitor by site-directed mutagenesis and chimerical protein studies can be achieved by modifications in the (poly)peptide primary sequence that affect the antagonist's or inhibitor's affinity; this usually allows to precisely map the binding pocket for the drug. Identification of binding sites may be assisted by computer programs. Thus, appropriate computer programs can be used for the identification of interactive sites of a putative antagonist or inhibitor and the polypeptide of the invention by computer assisted searches for complementary structural motifs (Fassina, Immunomethods 5 (1994), 114-120).

As regards step (b), the following protocols may be envisaged: Once the effector site for antagonists or inhibitors has been mapped, the precise residues interacting with different parts of the antagonists or inhibitors can be identified by combination of the information obtained from mutagenesis studies (step (a)) and computer simulations of the structure of the binding site provided that the precise three-dimensional structure of the antagonists or inhibitors is known (if not, it can be predicted by computational simulation). If said antagonist or inhibitor is itself a peptide, it can be also mutated to determine which residues interact with others in the above-mentioned polypeptide essential for bacterial growth and survival.

Finally, in step (c) the antagonist or inhibitor can be modified to improve its binding affinity or its potency and specificity. If, for instance, there are electrostatic interactions between a particular residue of an polypeptide as defined above and some region of an antagonist or inhibitor molecule, the overall charge in that region can be modified to increase that particular interaction. Furthermore, the three-dimensional and/or crystallographic structure of inhibitors or antagonists of the polypeptide of the invention can be used for the design of peptidomimetic inhibitors or antagonists, e.g. in combination with said polypeptide (Rose, Biochemistry 35 (1996), 12933-12944; Rutenber, Bioorg. Med. Chem. 4 (1996), 1545-1558).

Potential antagonists/inhibitors include antisense molecules. Antisense technology can be used to control gene expression through antisense DNA or through triple-helix formation. Antisense techniques are discussed, for example, in Okano, J. Neurochem. 56 (1991), 560; Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988). Triple helix formation is discussed in, for instance, Lee, Nucl. Acids Res. 6 (1979), 3073; Cooney, Science 241 (1988), 456; and Dervan, Science 251 (1991), 1360. The methods are based on binding of a polynucleotide to a complementary DNA or RNA.

For example, the 5' coding portion of a polynucleotide that encodes the mature polypeptide as described above may be used to design an antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription thereby preventing transcription and the production of the protein. The antisense RNA oligonucleotide hybridizes to the mRNA and blocks translation of the mRNA molecule into receptor polypeptide. As indicated, antagonist or inhibitor e.g. polyclonal and monoclonal antibody according to the teachings of the present invention can be raised according to the methods disclosed in Tartaglia, J. Biol. Chem. 267 (1992), 4304-4307; Tartaglia, Cell 73 (1993), 213-216, and PCT Application WO 94/09137.

Antibodies may be prepared by any of a variety of methods using immunogens of the polypeptide described above. As indicated, such immunogens include the full length polypeptide (which may or may not include the leader sequence) and fragments such as the ligand binding domain, the extracellular domain and the intracellular domain. These antibodies can be monoclonal antibodies, polyclonal antibodies or synthetic antibodies as well as fragments of antibodies, such as Fab<sup>+</sup>, Fv, F(ab')<sub>2</sub>, disulphide-bridged Fv or scFv fragments, etc. Monoclonal antibodies can be prepared, for example, by the techniques as originally described in Köhler and Milstein, Nature 256 (1975), 495, and Galfré, Meth. Enzymol. 73 (1981), 3, which comprise the fusion of mouse myeloma cells to spleen cells derived from immunized mammals. Furthermore, antibodies or fragments thereof to the aforementioned peptides can be obtained by using methods which are described, e.g., in Harlow and Lane "Antibodies, A Laboratory Manual", CSH Press, Cold Spring Harbor, 1988.

The antagonists or inhibitors isolated by the above methods also serve as lead compounds for the development of analog compounds. The analogs should have a stabilized electronic configuration and molecular conformation that allows key functional groups to be presented to the receptor in substantially the same way as the lead compound. In particular, the analog compounds have spatial electronic properties which are comparable to the binding region, but can be smaller molecules than the lead compound, frequently having a molecular weight below about 2 kD and preferably below about 1 kD. Identification of analog compounds can be performed through use of techniques such as self-consistent field (SCF) analysis, configuration interaction (CI) analysis, and normal mode dynamics analysis. Computer programs for implementing these techniques are available; e.g., Rein, Computer-Assisted Modeling of Receptor-Ligand Interactions (Alan Liss, New York, 1989). Methods for the preparation of chemical derivatives and analogues are well known to those skilled in the art and are described in, for example, Beilstein, Handbook of Organic Chemistry, Springer edition New York Inc., 175 Fifth Avenue, New York, N.Y. 10010 U.S.A. and Organic Synthesis, Wiley, New York, USA. Furthermore, said derivatives and analogues can be tested for their effects according to methods known in the art. Furthermore, peptidomimetics and/or computer aided design of appropriate derivatives and analogues can be used, for example, according to the methods described above.

The inhibitor or antagonist identified by the above-described method may prove useful as a pesticide, and/or antibiotic. The inhibitors and antagonists of the present invention preferably have a specificity at least substantially identical to the binding specificity of the natural ligand or binding partner of the polypeptide described above. An antagonist or inhibitor can have a binding affinity to said polypeptide of at least  $10^5 M^{-1}$ , preferably higher than  $10^7 M^{-1}$  and advantageously up to  $10^{10} M^{-1}$ . In a preferred embodiment, an inhibitor, e.g. suppressive antibody, has an affinity of at least about  $10^{-7} M$ , preferably at least about  $10^{-9} M$  and most preferably at least about  $10^{-11} M$ ; and the antagonist has an affinity of less than about  $10^{-7} M$ , preferably less than about  $10^{-9} M$  and most preferably in order of  $10^{-11} M$ .

In the case of nucleic acid molecules it is preferred that they have a binding affinity to those encoding the amino acid sequences encoded in any one of SEQ ID NOS: 16 to 39 of at most 2-, 5- or 10-fold less than an exact complement of 20 consecutive nucleotides of the above described nucleic acid molecules.

In another embodiment, the present invention relates to a method for producing a therapeutic agent comprising synthesizing the above-described antagonist or inhibitor.

Preferably, the compound identified according to the above described method or its analog or derivative is further formulated in a therapeutically active form or in a form suitable for the application against bacterial infections or diseases related to such an infection. For example, it can be combined with a pharmaceutically acceptable carrier known in the art. Thus, the present invention also relates to a method of producing a (therapeutically effective) composition comprising the steps of one of the above described methods of the invention and combining the compound obtained or identified in the method of the invention or an analog or derivative thereof with a pharmaceutically acceptable carrier.

Also, the present invention relates to a composition comprising the antagonist or inhibitor mentioned above. As is evident from the above, the present invention generally relates to compositions comprising at least one of the aforementioned antagonists or inhibitors, which may be nucleic acid molecules, proteins or antibodies. Advantageously, said composition is for use as a medicament, a diagnostic means, or a kit.

The term "composition", as used in accordance with the present invention, comprises at least one small molecule or molecule as identified herein above, such as a protein, an antigenic fragment of said protein, a fusion protein, a nucleic acid molecule and/or an antibody as described above and, optionally, further molecules, either alone or in combination, like e.g. molecules which are capable of optimizing antigen processing, cytokines, immunoglobulins, lymphokines or CpG-containing DNA stretches or, optionally, adjuvants. The composition may be in solid, liquid or gaseous form and may be, inter alia, in form of (a) powder(s), (a)

tablet(s), (a) solution(s) or (an) aerosol(s). In a preferred embodiment, said composition comprises at least two, preferably three, more preferably four, most preferably five differentially synthesized proteins.

The antagonists and inhibitors of the invention appear to function against gene products which are essential in several strains or genera of bacteria. Accordingly, the above-described antagonists and inhibitors may be used to inhibit the growth of a wide spectrum of bacteria. The above described antagonists or inhibitors may be used to slow, stop, or reverse bacterial growth. Thus, the present invention also relates to a method of producing a therapeutic agent comprising the steps of the methods described hereinbefore and synthesizing the antagonist or inhibitor obtained or identified as described above or an analog or derivative thereof, preferably in an amount sufficient to provide said agent in a therapeutically effective amount to a patient.

Compounds identified by the above methods or analogs are formulated for therapeutic use as pharmaceutical compositions. The compositions can also include, depending on the formulation desired, pharmaceutically acceptable, usually sterile, non-toxic carriers or diluents, which are defined as vehicles commonly used to formulate pharmaceutical compositions for animal or human administration. The diluent is selected so as not to affect the biological activity of the combination. Examples of such diluents are distilled water, physiological saline, Ringer's solutions, dextrose solution, and Hank's solution. In addition, the pharmaceutical composition or formulation may also include other carriers, adjuvants, or nontoxic, nontherapeutic, nonimmunogenic stabilizers and the like. A therapeutically effective dose refers to that amount of protein or its antibodies, antagonists, or inhibitors which ameliorate the symptoms or condition. Therapeutic efficacy and toxicity of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) and LD<sub>50</sub> (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD<sub>50</sub>/ED<sub>50</sub>.

Compositions comprising such carriers can be formulated by well known conventional methods. These pharmaceutical compositions can be administered

to the subject at a suitable dose. Administration of the suitable compositions may be effected by different ways, e.g., by intravenous, intraperitoneal, subcutaneous, intramuscular, topical, intradermal, intranasal or intrabronchial administration. The dosage regimen will be determined by the attending physician and clinical factors. As is well known in the medical arts, dosages for any one patient depends upon many factors, including the patient's size, body surface area, age, the particular compound to be administered, sex, time and route of administration, general health, and other drugs being administered concurrently. Proteinaceous pharmaceutically active matter may be present in amounts between 1 ng and 10 mg per dose; however, doses below or above this exemplary range are envisioned, especially considering the aforementioned factors. Administration of the suitable compositions may be effected by different ways, e.g., by intravenous, intraperitoneal, subcutaneous, intramuscular, topical or intradermal administration. If the regimen is a continuous infusion, it should also be in the range of 1 µg to 10 mg units per kilogram of body weight per minute, respectively. Progress can be monitored by periodic assessment. The compositions of the invention may be administered locally or systemically. Administration will generally be parenterally, e.g., intravenously. The compositions of the invention may also be administered directly to the target site, e.g., by biostatic delivery to an internal or external target site or by catheter to a site in an artery. Preparations for parenteral administration include sterile aqueous or non-aqueous solutions, suspensions, and emulsions. Examples of non-aqueous solvents are propylene glycol, polyethylene glycol, vegetable oils such as olive oil, and injectable organic esters such as ethyl oleate. Aqueous carriers include water, alcoholic/aqueous solutions, emulsions or suspensions, including saline and buffered media. Parenteral vehicles include sodium chloride solution, Ringer's dextrose, dextrose and sodium chloride, lactated Ringer's, or fixed oils. Intravenous vehicles include fluid and nutrient replenishers, electrolyte replenishers (such as those based on Ringer's dextrose), and the like. Preservatives and other additives may also be present such as, for example, antimicrobials, anti-oxidants, chelating agents, and inert gases and the like. Furthermore, the pharmaceutical composition of the invention may comprise further agents such as interleukins, interferons and/or CpG-containing DNA stretches, depending on the intended use of the pharmaceutical composition.

20.

In another embodiment, the present invention relates to a kit comprising at least one of the aforementioned antagonists or inhibitors of the invention. The kit of the invention as well as the composition may in a preferred embodiment contain further ingredients such as selection markers, antibiotics, cytokines and components for simplifying or supporting the treatment of bacterial infections or disorders or diseases related to bacterial infections. The kit of the invention may advantageously be used for carrying out the method of the invention and could be, inter alia, employed in a variety of applications referred to herein, e.g., in the diagnostic field or as research tool. The parts of the kit of the invention can be packaged individually in vials or in combination in containers or multicontainer units. Manufacture of the kit follows preferably standard procedures which are known to the person skilled in the art. The kit or its ingredients according to the invention can be used in antibacterial therapies, for example, for any of the above described methods for detecting further inhibitors and antagonists essential for bacterial growth and survival. The kit of the invention and its ingredients are expected to be very useful for the healing and protection of animals and humans suffering from a bacterial infection.

The present invention also relates to a method for treating or preventing bacterial infections or diseases or disorders related to bacterial infections comprising the step of administering to a subject in need thereof an antagonist or inhibitor identified herein above, optionally comprised in a pharmaceutical composition of the invention.

In another embodiment the present invention relates to the use of a polypeptide encoded by the gene as identified above or a fragment, derivative or ortholog thereof or of any of said genes for the identification of an antagonist or inhibitor of said polypeptide fragment, derivate or ortholog or said gene.

In a further embodiment the present invention relates to the use of said polypeptide, the therapeutic agent produced according to the invention, the antagonist or inhibitor obtained or identified by the method or use according to the invention for the preparation of a pharmaceutical composition for the treatment of

(a) bacterial infection(s), disorder(s) and/or disease(s) related to bacterial infections.

In another embodiment the present invention relates to a method for treating or preventing bacterial infections or diseases or disorders related to bacterial infections comprising the step of administering to a subject in need thereof an antagonist or inhibitor identified herein above, optionally comprised in the pharmaceutical composition according to the present invention.

In a further embodiment the present invention relates to the use of the above-described polypeptide, a fragment, derivative or ortholog thereof or of any of said genes for screening for polypeptides interacting with said polypeptide using protein-protein interaction technologies, and/or for validating such interaction as being essential for bacterial survival and/or for screening for antagonists or inhibitors of such interaction.

In a further embodiment the present invention relates to the use of the above-described polypeptide, a fragment, derivative or ortholog thereof or of any of said genes for screening of polypeptide for polypeptide binding to said polypeptide, and/or for validating the peptides binding to said polypeptide as preventing growth of bacteria or being lethal to bacteria upon expression of said polypeptides in said bacteria, and/or for screening for small molecules competitively displacing said peptides.

In another embodiment the present invention relates to the use of a conditional mutant of a gene as described above or a fragment, derivative or ortholog thereof or of surrogate ligands against said gene expressed in bacteria to induce a lethal phenotype in bacteria and/or for the analysis of said bacteria for surrogate markers by comparison of RNA or protein profiles in said bacteria with RNA or protein profiles in wild type bacteria, and/or the use of said surrogate markers for the identification of antagonists of the essential function of said gene.

In another embodiment the present invention relates to a method for identifying or isolating a surrogate marker comprising the steps as described in the above-recited method of the present invention.

In a further embodiment the present invention relates to a method for identifying or isolating a surrogate marker comprising the steps of

- (a) inducing a lethal phenotype in bacteria representing a conditional mutant of a gene selected from the group consisting of ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC; and
- (b) analysing said bacteria comparing the RNA or protein profile of said bacteria with wild type bacteria.

The invention also relates to the above recited genes and polypeptides and fragments, derivatives and orthologs thereof.

These and other embodiments are disclosed and encompassed by the description and examples of the present invention. Further literature concerning any one of the methods, uses and compounds to be employed in accordance with the present invention may be retrieved from public libraries, using for example electronic devices. For example the public database "Medline" may be utilized which is available on the Internet, for example under <http://www.ncbi.nlm.nih.gov/PubMed/medline.html>. Further databases and addresses, such as <http://www.ncbi.nlm.nih.gov/>, <http://www.infobiogen.fr/>, [http://www.fmi.ch/biology/research\\_tools.html](http://www.fmi.ch/biology/research_tools.html), <http://www.tigr.org/>, are known to the person skilled in the art and can also be obtained using, e.g., <http://www.lycos.com>. An overview of patent information in biotechnology and a survey of relevant sources of patent information useful for retrospective searching and for current awareness is given in Berks, TIBTECH 12 (1994), 352-364.

The present invention is further illustrated by reference to the following non-limiting examples.

Unless stated otherwise in the examples, all recombinant DNA techniques are performed according to protocols as described in Sambrook et al. (1989), Molecular Cloning : A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY or in Volumes 1 and 2 of Ausubel et al. (1994), Current Protocols in Molecular Biology, Current Protocols. Standard materials and methods for plant molecular work are described in Plant Molecular Biology Labfase (1993) by R.D.D. Croy, jointly published by BIOS Scientific Publications Ltd. (UK) and Blackwell Scientific Publications (UK).

#### Brief description of the figures

- Figure 1: Sequences of the essential bacterial genes identified according to the method described in the examples
- Figure 2: PCR strategy and the position of primers used
- Figure 3: Sequence comparison table of essential *E.coli* genes with proposed orthologs from various bacteria. Unfinished genomes are indicated by asterisk. Complete genomes were analysed using BlastP2. Unfinished genomes were analysed with TBLASTN. Orthologous sequences can be accessed at the respective WWW links as indicated in the footnotes.
- Figure 4: Multiple Sequence Alignment (MSA) of *E. coli* gene *ygbB* with orthologs in 5 different bacterial organisms including homology score. Similar MSA with similar results have been created for all 22 essential bacterial genes.

#### Example 1

An automated BLASTP-based genome comparisons to identify *E. coli* FUN genes resulted in the following list of 65 candidate genes which are conserved between *E. coli*, *B. subtilis*, *H. influenzae*, *H. pylori*, *M. tuberculosis*, *Ch. trachomatis*, *B.*

*burgdorferi*, *T. pallidum*, *S. pneumoniae*, *S. aureus*, *E. faecalis*, *P. aeruginosa*, *B. pertussis* and which were further analysed:

FUN Genes	Gene Bank Accession Number	FUN Genes	Gene Bank Accession Number	FUN Genes	Gene Bank Accession Number
ygbB	g1789103	yggS	g1789321	yaeE	g1786397
yhaD	g1789512	yggV	g1789324	yicC	g1790075
yhbU	g1789548	yggW	g1789325	yebK	g1788159
yhiN	g2367234	yjhG	g2367371	yhbC	g1789561
yieG	g1790150	yjiR	g1790797	ygbP	g1789104
yihZ	g1790320	yohI	g1788462	ybaX	g1786648
yjgF	g1790691	yqhThom	g1788728	yqcD	g1789158
yacE	g1786292	yfiH	g1788945	ybeY	g1786880
yaeC	g1786396	yhaR	g1789501	gcpE	g1788863
yagF	g1786464	yhdG	g1789660	kdtB	g1790065
ybeB	g1786856	yccG	g1787197	pfs	g1786354
ycfH	g147382	ychB	g1787459	sms	g1790850
ydcP	g1787705	yejD	g1788510	ycaJ	g1787119
ydiB	g1787983	yidD	g140861	yhhF	g1789875
yebI	g1788166	yrfl	g1789804	yleA	g1786882
yeeC	g1788320	yggJ	g1789315	b1808	g1788110

FUN Genes	Gene Bank Accession Number	FUN Genes	Gene Bank Accession Number	FUN Genes	Gene Bank Accession Number
yegQ	g1788397	yjeE	g1790610	yeaA	g1788077
yfcB	g1788670	yiaO	g1790004	b1675	g1787964
yfgB	g1788865	yrdC	g2367210	yhbU/yegQ <sup>a</sup>	g1789548 / g1788397
yfhC	g1788911	b1983	g1788294	yjgF/yhaR <sup>a</sup>	g1790691 / g1789501
ydiD	g1787993	yeeS	g1736671	b2385	g1788728
nlpA	g72589	yaaJ	g1786188	yicO	g1790097
yfjY	g1788997	ydhE	g1742737	yebC	g140614
ykfG	g2367100	yjcD	g396399	yohI/yhdG <sup>a</sup>	g1788462 / g1789660
ygcA	g1789148	yceG	g1787339	smpB	g1788973
ygfA	g1789278	yjbC	g396357		

<sup>a</sup>: double mutants were created when the respective genes were paralogues in *E.coli*

#### Creating in-frame deletions of *E. coli* genes

The subsequent description of the construction of deletion mutants was carried out essentially equal for these 77 candidate genes. Particular details will exemplarily be described for one gene which gave rise to be essential (yfhC) and one which was non-essential (yggV).

**1) Principle of the PCR-procedure and primer-design for *in frame* deletions:**

Unless an overlapping ORF exists, primers dgenX2 and dgenX3 are designed to delete the entire ORF from ATG to STOP, e.g.: ATGgttataaaattggaggtgtgaagggtattgcgtgTAA (SEQ ID NO: 1) (see figure). The 5'-ends of primers dgenX1 and dgenX4 contain random nucleotides followed preferably by a BamHI site (dgenX1) or a Sall site (dgenX4) for cloning into plasmid pKO3 (Link et al (1997), J Bac 179: 6228-6237). In most mutants, primers dgenX2 and dgenX3 contain a 33 bp tag sequence called "Church-tag".

Church-tag forward direction: 5'-gttataaaattggaggtgtgaagggtattgcgtg-3' (SEQ ID NO: 2)

Church-tag reverse direction: 5'-cacgcaataacccatcacactccaaattataac-3' (SEQ ID NO: 3)

This tag is used for a subsequent PCR in which the 5'- and 3'- flanking DNA-fragments of the deletion construct are assembled.

In the few constructs lacking the "Church-tag", the primers dgenX2 and dgenX3 carry at their 5'-ends 5 random nucleotides followed by a restriction site (preferably EcoRI) which by its positioning creates the *in frame* deletion.

Oligos cgenX1 and cgenX2 are used for the verification of the chromosomal situation (wild type or deletion) after the replacement procedure (Fig. 2).

Primers for the respective candidate genes were designed as follows:

dyfhC1: 5'-GATCGGATCCAAATTCCAGTTAGCCATGATGCGGTC-3'  
(SEQ ID NO: 4)

dyfhC2: 5'-CACGCAATAACCTTCACACTCCAAATTATAACCATTATA  
CACGGACGCTATGC-3' (SEQ ID NO: 5)

dyfhC3: 5'-  
GTTATAAATTGGAGTGTGAAGGTTATTGCGTGACGGATTAATT  
TTGTTTCTCTT-3' (SEQ ID NO: 6)

dyfhC4: 5'-GATCGTCGACGCGCTCGATATCACCGATGAACAAACCG-3'  
(SEQ ID NO: 7)

cyfhC1: 5'-CAATCCGCTGCTTATTCTGTCAG-3' (SEQ ID NO: 8)

27

cyfhC2: 5'-TTATAACGAAATCAACGGGAAACCT-3' (SEQ ID NO: 9)

dyggV1: 5'-GATCGGATCCCTCTAAAAAATAAGGAATTAAAGG-3'  
(SEQ ID NO: 10)

dyggV2: 5'-CACGCAATAACCTTACACTCCAAATTATAACCATAGGATAC  
CTAATTAATTAAC-3' (SEQ ID NO: 11)

dyggV3: 5'-GTTATAAAATTGGAGTGTGAAGGTTATTGCGTGAAGAGCGCC  
ATTCCCACCGT-3' (SEQ ID NO: 12)

dyggV4: 5'-GATCGTCGACTCATATTGCTGATAACCCGCTGCGGT-3'  
(SEQ ID NO: 13)

cyggV1: 5'-GTTGACGGCCAGGCCAACAGTCAT-3' (SEQ ID NO: 14)

cyggV2: 5'-ATAACCTGGCAATCGCCTCG-3' (SEQ ID NO: 15)

## Example 2

### Construction of the DNA-fragments comprising the deletion

The 5'- and the 3'-flanking DNA fragments are PCR amplified in a total volume of 50 µl as follows:

Chromosomal DNA from *E. coli* strain MG1655 (100 ng/µl):

	final conc.: 1 ng/µl
10*Pwo-buffer	final conc.: 1x
dgenX1/3 (10 µM)	final conc.: 500 nM
dgenX2 (4) (10 µM)	final conc.: 500 nM
Pwo-Polymerase	final conc.: 5 U/100 µl
dNTPs (25 mM)	final conc.: 250 µM
H <sub>2</sub> O	to adjust volume to 50 µl

### PCR conditions:

4' 94 °C

28

30 cycles: 30" 94 °C, 30" 44 °C, 1' 72 °C

5' 72 °C

The PCR products are then purified with the High Pure PCR Purification Kit (Boehringer) to remove salts and enzyme (elute in 50 µl H<sub>2</sub>O). Alternatively, if PCR products contain prominent impurities, the respective fragment must be purified by agarose gel extraction (Gene Clean, Dianova) before the fragment assembly.

#### Assembly PCR

Equal amounts of 5'- and 3'-fragment are applied as template DNA. In general a volume applied for gel electrophoresis giving an intense band is o.k. The total reaction volume is 100 µl. For the assembly the "outer" primers dgenX1 and dgenX4 were used.

5'-Fragment	approx. 10 ng
3'-Fragment	approx. 10 ng
10*Pwo-buffer	final conc.: 1x
dgenX1 (10 µM)	final conc.: 500 nM (50 pmol/100 µl)
dgenX4 (10 µM)	final conc.: 500 nM (50 pmol/100 µl)
Pwo-Pol (Boehringer)	final conc.: 5 U/100 µl
dNTPs (25mM)	final conc.: 250 µM
H <sub>2</sub> O	add to 100 µl

#### PCR conditions:

4' 94 °C

10 cycles: 30" 94 °C, 30" 44 °C, 1' 72 °C

25 cycles: 30" 94 °C, 30" 44 °C, 3' 72 °C

5' 72 °C

The success of the PCR is checked by agarose gel electrophoresis. The assembled PCR product is purified with the High Pure PCR Purification Kit and the complete eluate of 50 µl is over-night digested with BamHI and Sall in a volume of 60 µl. After gel electrophoresis the digested product is purified with

5, 65 °C

25 cycles: 30° 94 °C, 30° 50 °C, 2° 65 °C

4, 94 °C

**PCR conditions:**

H <sub>2</sub> O	15.35 μl
dNTPs (25 mM)	final conc.: 250 μM
Taq-Pol (QIAGEN)	final conc.: 2 U/25 μl
PKO3-S1/dgenX4 (100 μM)	final conc.: 1 μM (50 pmol/100 μl)
PKO3-B1/dgenX1 (100 μM)	final conc.: 1 μM (50 pmol/100 μl)
5*Q-solution	final conc.: 1x
10*Taq-buffer	final conc.: 1x
template (colony)	1 μl of 1 colony resuspended in 20 μl H <sub>2</sub> O

**Reaction mixture for 25 μl reaction volume:**

gene specific primers (dgenX1 and dgenX4).

2) Clones with the correct size of insert are double-checked by colony-PCR with (PKO3-B1 and PKO3-S1).

1) 8 clones are characterized by colony-PCR with vector PKO3-specific primers

**Verification of deletion constructs:**

One half of the ligation mix is transformed into chemically competent E. coli DH5α and clones are purified once (usually 8 clones are sufficient).

**Transformation into DH5:**

10-20 μl reaction (T4-DNA ligase) for 2 hours at room temperature.

Next, the fragment is ligated into the vector PKO3 (cut with BamHI and SalI) in a

**Cloning into vector PKO3:**

volume: 25 μl).

Gene Clean (Dianova) to remove small oligonucleotides quantitatively (elution 29

30

- 3) Plasmid-DNA from 4 ml over-night culture is prepared using a QIAgen Miniprep Kit and a double restriction analysis with BamHI/Sall and EcoRI/HindIII is performed to verify the clones.

**Protocol referring to the construction of assembly products by a restriction site:**

The 5'- and the 3'-fragments are PCR amplified as described above. The PCR products are purified with the High Pure PCR Purification Kit (Boehringer) to remove salts and enzyme and 5 to 10 µl are digested over night using the restriction site creating the deletion (primers 2 and 3; mostly EcoRI) in a total volume of 30 µl. The restriction products are again purified with the High Pure PCR Purification Kit to remove nucleotides, salts and enzyme. (Alternatively: Following preparative agarose gel electrophoresis the cut fragments are isolated using Gene Clean (Dianova) and eluted in a volume of 25 µl. The cut fragments (3-6 µl each) are ligated in a volume of 10-15 µl using T4-DNA ligase for 2 hours at room temperature. 5 µl of this ligation mix is directly used as a template for a second PCR. In this PCR, the assembled fragments are amplified using primers dgenX1 and dgenX4. The reaction is set up as described above with two exceptions: 1) The total reaction volume is 100 µl and 2) the extension step at 72 °C lasts 3'.

**Example 3**

**The chromosomal exchange strategy**

(Link et al (1997), J Bac 179: 6228-6237)

**Cointegration:**

Cointegration = integration of a plasmid into the chromosome by a recombination event

The pKO3 derivative is transformed into MG1655 or any recA+ strain

**Day 1**

The strain is grown at 30 °C in LB containing 20 µg/ml chloramphenicol (LB-Cam20) to an OD<sub>600</sub> of ~1.0. Afterwards, perform 10-fold serial dilutions in the same medium (down to 10<sup>-7</sup>). For the following plating use prewarmed LB-Cam20 agar plates. Plate 100 µl of dilutions 10<sup>-4</sup> and 10<sup>-5</sup> for incubation at 44 °C and 100 µl of dilutions 10<sup>-6</sup> and 10<sup>-7</sup> for incubation at 30 °C.

**Day 2**

Following incubation at the respective temperature, determine the factor c.f.u.44 °C/c.f.u.30 °C (c.f.u. = colony forming units). This factor for pKO3 without insert is in the range 1\*10<sup>-4</sup> to 5\*10<sup>-4</sup> and should be significantly larger in the case of successful cointegration. Purify 8 randomly chosen clones from the 44 °C plate twice on LB-Cam20 agar plates at 44 °C (during Day 2 and over night to Day 3). Optionally, confirm the clones for their identity as cointegrates by colony-PCR.

**Resolution and counter-selection:**

Resolution = resolution of the cointegrate resulting in a self replicative plasmid by a second recombination event

Counter-selection = selection against the presence of plasmid in the cell

**Day 3**

Pool single colonies from each of the 8 cointegrates in 100 µl LB and use this suspension as an inoculum for 10 ml LB+5 % sucrose. After growth at 30 °C (8 to 10 hours during a day is sufficient) 10-fold serial dilutions are performed and 100 µl of dilutions 10<sup>-4</sup>, 10<sup>-5</sup>, and 10<sup>-6</sup> are plated onto LB agar+5 % sucrose and grown over night at 30 °C.

**Day 4**

50 single colonies are replica streaked on LB+Cam20 and LB+5 % sucrose to test for the loss of plasmid.

32  
Example 4**Testing for essentiality of FUN genes of *E. coli* and interpretation of the results****Day 5**

The clones sensitive to chloramphenicol are then tested for their genotype (wild type versus in-frame deletion) by colony-PCR using primers cgenX1 and cgenX2 (10-48 clones).

In the case of the gene *yfhC* out of 48 clones tested only wild type situation on the chromosome could be detected.

In the case of the gene *yggV* out of 48 clones 16 (= 33 %) revealed a PCR product with a size indicative for the deletion situation on the chromosome.

Are 48 clones revealing no mutant enough to claim a gene as essential? This question can be answered by asking for the number of clones that have to be tested to get a confidence of e.g. 99 % that really no mutants are present in an infinite number of clones. Provided a hypothesis  $H_0$  means that only the wild type genotype is viable and hypothesis  $H_1$  means that a fraction ( $1-x$ ) of mutants is allowed to occur together with the wild type ( $x$ ) among a population of clones ( $x + (1-x)$ ), then the probability to make the wrong decision (decision for  $H_0$  whereas  $H_1$  is true) can be calculated as

$$(1) \quad x^n / (1+x^n)$$

where  $x$  is the fraction of wild type clones and  $n$  is the number of clones tested.

The confidence niveau  $\alpha$  to make the wrong decision (error probability) is given by

$$(2) \quad \alpha > x^n / (1+x^n)$$

thereby resulting in

$$(3) \quad n > \ln(\alpha / (1-\alpha)) / \ln(x)$$

for the number of clones that have to be tested to prove or disprove hypothesis  $H_0$ .

33

If the average probability for obtaining wild type clones ( $x$ ) in a replacement experiment is 70 % (experimentally determined for 43 non-essential genes out of 65 candidate genes), then, after testing of 26 clones which reveal a wild type genotype an uncertainty of 0.01 % error probability ( $\alpha$ ) remains that the claiming of a gene as essential could be wrong. Even if the rate of obtaining wild types ( $x$ ) is set to 85 % (a value which occurs with a frequency of 10 % for replacement experiments with non-essential genes), then, by testing 32 clones (which was performed in every experiment giving rise to an essential gene) an error probability of only 0.6 % remains to chose the wrong hypothesis.

### Example 5

#### List of essential FUN genes obtained

By the described method the following 24 genes were obtained which gave no deletion genotype and are therefore claimed to be essential:

<i>E. coli</i>	
gene name	GenBank#
ygbB	g1789103
yfhC	g1788911
yacE	g1786292
ychB	g1787459
yejD	g1788510
yrfl	g1789804
yggJ	g1789315
yjeE	g1790610
yiaO	g1790004
yrdC	g2367210
yhbC	g1789561
ygbP	g1789104
ybeY	g1786880

gcpE	g1788863
kdtB	g1790065
pfs	g1786354
ycaJ	g1787119
b1808	g1788110
yeaA	g1788077
yagF	g1786464
b1983	g1788294
yidD	g140861
yceG	g1787339
yjbC	g396357

## CLAIMS

1. A method for identifying an antagonist or inhibitor of the expression of a gene encoding a polypeptide essential for bacterial growth or survival wherein said gene is selected from the group consisting of ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC, the sequence of said genes being shown in Fig. 1, or a fragment, derivative or ortholog thereof, said method comprising the steps of
  - (a) testing a candidate antagonist or inhibitor or a sample comprising a plurality of said candidate antagonists or inhibitors for the inhibition or reduction of transcription of said gene or a fragment or derivative thereof; or
  - (b) testing a candidate antagonist or inhibitor or a sample comprising a plurality of said candidate antagonists or inhibitors for the inhibition or reduction of translation of mRNA transcribed from said gene or a fragment or derivative thereof; and
  - (c) identifying an antagonist or inhibitor or a sample comprising a plurality of said candidate antagonists or inhibitors that tests positive in step (a) and/or (b).
2. A method for testing a candidate antagonist or inhibitor of a polypeptide or a mRNA essential for bacterial growth or survival encoded by a gene selected from the group consisting of ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC, the sequence of said genes being shown in Fig. 1, or a fragment, derivative or ortholog thereof comprising the steps of
  - (a) contacting a bacterial cell with a candidate antagonist or inhibitor or a sample comprising a plurality of said candidate antagonists or inhibitors; and
  - (b) testing whether said contacting leads to cell growth inhibition and/or cell death.

3. A method for testing a candidate antagonist or inhibitor of the function of a gene essential for bacterial growth or survival wherein said gene is selected from the group consisting of ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC, the sequence of said genes being shown in Fig. 1, or a fragment, derivative or ortholog thereof, comprising the steps of
  - (a) contacting a bacterial cell comprising said gene with a candidate antagonist or inhibitor or a sample comprising a plurality of said candidate antagonists or inhibitors; and
  - (b) testing whether said contacting leads to cell growth inhibition and/or cell death.
4. The method of any one of claims 1 to 3 further comprising identifying an antagonist or inhibitor, optionally from said sample of candidate antagonists or inhibitors.
5. The method of any one of claims 1 to 4 wherein said inhibitor or antagonist is further improved by peptidomimetics or by applying phage display or combinatorial library technique step(s).
6. A method for designing an improved antagonist or inhibitor for the treatment of a bacterial infection or disorder or disease related to a bacterial infection comprising the steps
  - (a) identification of the binding site of an antagonist or inhibitor to the polypeptide ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC, the sequence of said genes being shown in Fig. 1, or obtained by or identified by the method of any one of claims 1 to 5 by site-directed mutagenesis and chimeric polypeptide studies;
  - (b) molecular modeling of both the binding site of said antagonist or inhibitor and the structure of said polypeptide; and

- (c) modification of said antagonist or inhibitor to improve its binding specificity or affinity for the polypeptide.
7. An antagonist or inhibitor of the activity of a polypeptide encoded by a gene selected from the group consisting of ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC, the sequence of said genes being shown in Fig. 1, or fragment, derivative or ortholog thereof or of the expression of a gene encoding said polypeptide or said fragment, derivative or ortholog or obtained by or identified by the method of any one of claims 1 to 6.
8. A method for producing a therapeutic agent comprising synthesizing the antagonist or inhibitor identified, tested or designed according to the method of any one of claims 1 to 6 or the antagonist or inhibitor of claim 7 or an analog or derivative thereof.
9. A method for producing a composition comprising the steps of the method of any one of claims 1 to 6 or synthesizing the antagonist or inhibitor of claim 7 and formulating said inhibitor or antagonist in a pharmaceutically acceptable form.
10. A composition comprising an antagonist or inhibitor of claim 7, the therapeutic agent produced by the method of claim 8 or the antagonist or inhibitor obtained by or identified in the method of any one of claims 1 to 6 or produced according to claim 9 and optionally a pharmaceutically acceptable carrier.
11. The composition of claim 10 which is a pharmaceutical composition.
12. The composition of claim 10 which is a kit.
13. The composition of any one of claims 10 to 12 further comprising an antibiotic and/or cytokine.

14. Use of a polypeptide encoded by a gene selected from the group consisting of ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC, the sequence of said genes being shown in Fig. 1, or a fragment, derivative or ortholog thereof or of any of said genes for the identification of an antagonist or inhibitor of the activity of said polypeptide or said fragment, derivative or ortholog or of the expression of a gene encoding said polypeptide or said fragment, derivative or ortholog.
15. Use of an antagonist or inhibitor of claim 7, the therapeutic agent produced by the method of claim 8 or the antagonist or inhibitor obtained by or identified in the method of any one of claims 1 to 6 or produced according to claim 9 or identified by the use of any of the claims for the preparation of a pharmaceutical composition for the treatment of (a) bacterial infection(s), disorder(s) and/or disease(s) related to bacterial infections.
16. A method for treating or preventing bacterial infections or diseases or disorders related to bacterial infections comprising the step of administering to a subject in need thereof the antagonist or inhibitor obtained by or identified in the method of any one of claims 1 to 6 or produced according to claim 9 optionally comprised in the pharmaceutical composition according to claim 11.
17. Use of a polypeptide encoded by a gene selected from the group consisting of ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC, the sequence of said genes being shown in Fig. 1, or a fragment, derivative or ortholog thereof or any of said genes for screening for polypeptides interacting with said polypeptide using protein-protein interaction technologies, and/or for validating such interaction as being essential for bacterial survival and/or for screening for antagonists or inhibitors of such interaction.

18. Use of a polypeptide encoded by a gene selected from the group consisting of ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC, the sequence of said genes being shown in Fig. 1, or a fragment, derivative or ortholog thereof or any of said genes for screening of polypeptide for polypeptide binding to said polypeptide, and/or for validating the peptides binding to said polypeptide as preventing growth of bacteria or being lethal to bacteria upon expression of said polypeptides in said bacteria, and/or for screening for small molecules competitively displacing said peptides.
19. Use of conditional mutants in a gene selected from the group consisting of ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC, the sequence of said genes being shown in Fig. 1, or a fragment, derivative or ortholog thereof or of surrogate ligands against said gene expressed in bacteria to induce a lethal phenotype in bacteria and/or for the analysis of said bacteria for surrogate markers by comparison of RNA or protein profiles in said bacteria with RNA or protein profiles in wild type bacteria, and/or the use of said surrogate markers for the identification of antagonists of the essential function of said gene.
20. A method for identifying or isolating a surrogate marker comprising the steps as described in claim 19.
21. A method for identifying or isolating a surrogate marker comprising the steps of
  - (a) inducing a lethal phenotype in bacteria containing a conditional mutant of a gene selected from the group consisting of ygbB, yfhC, yacE, ychB, yejD, yrfl, yggJ, yjeE, yiaO, yrdC, yhbC, ygbP, ybeY, gcpE, kdtB, pfs, ycaJ, b1808, yeaA, yagF, b1983, yidD, yceG and/or yjbC, the sequence of said genes being shown in Fig. 1; and

40

- (b) analysing said bacteria comparing the RNA or protein profile of said bacteria with wild type bacteria.

1 / 15

Fig. 1

1) yqbB

atgcgaattggacacgggttgcgatcgccttgcggtaaggccaaattatcatggtggcgtagcattccctacgaaa  
aggattgtggcgcatctgtggcgacgtggcgccatcgcttgaccgatcgatgtggcgccggcgctggggatat  
cgcaagctgtcccgataccgcattaaaggcgatagccgcgagctacgcgaagcctggcgtgtatcc  
aggcgaagggttatacccttgcacgcgtcatgtactatcatgcgtcaggcaccgaagatgtggcgacatccacaaatg  
cgctgtttatggccaagatcggctgcataatggatgtatgttacgtgaaagccactactacggaaaaactggattaccg  
gacgtggggaaagggttgcgtgaagcggtggccactactcattaaqcaacaaaatqa (SEQ ID NO: 16)

2) yfhC

atgcggccgcgtttataaccggagttctttgtcgaaatcgacgaaatctggatgcgtacgcgtacgcgt  
gcgaaacgtgcgtggatgagcggaaagtgcggcgcggattagtgcataacaatcggtaatcggtacggct  
gaaccggcccgatgtgcgcattatcccacccgacatcgaaaatcatggccgcggcagggtgtctggatgc  
tatcgctgtatgcacgcacccgttgtatgtcgcgttgcaccatgttaatgtgtgcggagcgatcc  
acgtgcgtatggtcgcgttggtgcgcgtacgcgaaaactggcgcgtggatcittaaatggatgc  
gtgcgtatccggatgtgaatcaccgcgtggaaattacggaaatgcggatgaglgcgcggcgttgc  
catgtacttgcatgcgcgcaggaaattaaagcgcgaaaaaaagcgcataatcctgcacggattaa  
agcgcgaaaaaaagcgcataatcctgcacggattaa (SEQ ID NO: 17)

3) vacE

atagggtatatagtgccttaacggggcattggcagtggcaagactaccgttgcctatcgaaattaac  
tcattgtgccatattattgcgcgtcagggtgtaaccagggtgcacctgcgtacatgcctatgcgtacatttggcgtaaca  
tgattgtcgatgtggAACATTGCGCCGGGCGCTTGCAGCAGCGGATCTGCCAACCCGAAGAGAAAAACTGGCTAA  
CGCCCTGCTGATCCGCTGATTCAAGAGACGCAACACCCAGATCCAGCAAGCTACTCCCCATGTACTGTGGGTGTGCC  
ATTGCTGGTAGAAAACCTACTGTATAAAAAAGCGAATCGAGTCAGTGTGGATGTCAGCCCAGAAACGCAACTTAAGCGC  
ACCATGCGAGCGCATGATGTAACCTCGCGAGCATGTCGAACAAACCTGCTGTCAGGCAACCGCGCAAGCCCGCTTGC  
TGGCAGATGACGTCAATTGATAATAACGGCGCACCGGATGCTATCGCATGGATGTTGCCGCTGCACGCACACTATTGCG  
TTCGCTGCAGTTGTCACAGGAAAAACCGTAA (SEQ ID NO: 18)

4) ychB

5) veiD

atgcgactgtaaaatttatcgcacagaactcgccgttagccgttattccggcgtaaatccggcaatcggtcacc  
gtcgatggcaaattcgccgtaatcgccgttcaactgtttctgaacatgtatgcgttacgtggcaacccgtggcag  
caacacgggtccacgttacttcatgctcaataaggctcagggtatgtttgtccacggacgaccctgtatccccaaacgggtctt  
atttcttgtaaaccggtagcgtggaaactgcgtggccgggggggttggatattgtataccaccggctgttgatgactga  
tgatggtcagtgttcgcaccgcattacttccgcgcatttgcgagaagacctatctgtgcactggaaatcacctgtact  
gacgataccgcagagcaattgtctaaaggcgtcagctgcataacgaaaaagatctactaaaggctcgggtctggaaatgt  
attaccccaacgcagggtcgtctgaccatcagcgaaggccgttacatcaggtggaaacgcgttccgcgcgtggtaacc  
cgtgggtggactgcgttgcacgttattccgggttacgcgttgcgtttagccccccgtgaatatcgccgttaacigaag  
aagaaaatggccagcgtcgtctaa (SEQ ID NO: 20)

6) yrfi

atgattatggcgaacatgaccaattacatcgctatctgttggaaaacttgcgtgcgcggcgactggtaaccgtttcgaaaa  
ccctgcacacagatccgtggaaaccacgattatccgcagccgttaaaaacgtgtggcagaactgctggcgaccagct  
gttaaccgcgtacgcgtggaaatgtgtatcacgcacagctgcaggcgacgggtccgtatggcggttattaaacggt  
aacaataaccaggcggatgcgggtggcgccgtgcaggcgaaatccaggaaaatgcgcacctgaaaacgcgtgtcg  
gcaatgggtacgtgtgtatcaccattaccccgagcgaaggcgaacgcatacaggcgtagtgtgtctggaaagggtatcccc  
gcggcctgcctggaaagattactttatgcgttcgaacagctgcgcacgcgcgtttattcgcacccggcgacgtacggc  
ccggcgtcaggcggtatgtgtgcaggtaatgcgtcgcaaaatgcgcacctggcgtacgcacttgc  
ccggaaaccatcaaaccggaaagaactgcgtacccgcacacgcggaaacgcggaaatgcgtgtggcggttgcacgc  
cggttacgatccgcaggatgtggagtcaaatgcacccgcgtgcgtgcacgcgtgcgtggcgatgcgtgaaaacgcgtgc  
aagaagtgtatgcatccgtggcgaaatggcgaaatgcacgcgtgtgttactgcgtgttgcgttgcgttgc  
ggatattgcgtggaaatccgcacacgcgtctccggcagatccgcacgcgttgcgttgcgttgcgttgcgttgc  
(SEQ ID NO: 21)

7) yggJ

8) yjeE

atgatgaatcgagtaattccgtccccgtatgagcaggcaacattagacctggggcagcgggtacgcgaaagccctgcgtggc  
gcaaccgttaatctatctgtatggcgatttaggcgcaggtaaaaccaccttttagccgggcttttacaggctctgggtcatcagg  
gtaatgcaaaaagccccacttatacgcgttgtcgAACCTatacgcgtcgacaacttaatggctatcacittgtatccgcctt  
ccgtatcccggaggagctggagttatggggatccgcgttttgcacaacgtgcacatctgcgttgggagtggccacaacaa  
ggtagcagggtgtctccgtacccggatgtcgaaatacacatgtatttcaggcacaaggccgtgaggcgcgcgtgagtcgggt  
tccctgcgggtgaattgtgtggcgcttagccggtaa  
(SEQ ID NO: 23)

9) viaO

atggaaattacgctgttaacctacgcatttcatgtggccgtgcattcgacacatcttcgtggcggcacaatcttacgtt  
cggttatgaacatcacaaaaccgactgcaacataatggcgcaaaaaattcaatgattttatgcaggagagaaccggc  
gagcgaaataaaactgttcccgacagcacttcgttaacgcgcaggcgatgtacgcggcgatgtccgttccgcaccatc  
gatatggaaatgtccggctcgaaataacttgcgggttatcaccagtgtaaacttgcgtatgtcccttcgttccgcataccg  
ctcacgcgcataaaacgcgtcgacggcaaagtgcgtatgtacccgttccgttccgttccgcataccg  
cctactggaaaaacggctggcgatgtcaccactgcgcgcacccgttaaaaaaccccgccgaccctgtaaaggactgaaaggactgaa  
atccgcaccaacaatagcccgtatgcgcattcaaaagtcttgcgtactcccgatccgtccgttccgcacgt  
tataccgggctggaaaccgcactatgcgcgtcaggAACACCCGATCAACGTCTGGTCAAGAAATTTCGAAGTGC  
GAAGTTCTCTGACGCACCGCCTATCCCCGCTTGGTGTATCAACAAAGCGAAGTTGATGGCTTAAGTCCGGAGGT  
CCAGCAGGCCTAGTTCTGACCAAGAAGCGGGTAACTATCAGCGCAAACGGTGTCAAGATCAGCAAAAAATCATC  
GACGGCATGAAAGAAGCGGGCGTGGAGTCATCACCGATCTGACCGCAAAGCCTTAGCGACGCACTGGGAATCAGGT  
TCGCGACATGTTGTTAAAGATGTGCCGAGGGAGCTGATCTGCTGAAAGCCGTGGATGAGGTGCAATAA (SEQ ID NO:  
24)

10) yrdC

cttgctgtacgagtcaccgaccatccgttgttgcgttgcgcaggctatggtaaaaccgcgtggttctaccagtgcacaacttga  
gtggattgcaccctgtcaaacagtagacgcagaagtgcgcacaattttggcgccccgttcccggttgccttgtgaaacgggg  
gggcgttaaatccctcagaaatccgcgtacccctgcacgggtgaactgtttcgcacaggggiaa (SEQ ID NO: 25)

11) yhbC  
gtgggcgttccacatttagagcaaaaattaacagagatgattactcgccagtgaggccctgggtttgaactggtgtcatcg  
aatttattcgcgtcgacatccacactgcgcatacttatattgtatagtaagatggcatcaatgtatgttgtatgtgacc  
accaggtaagtqtcgtqtcggatgttaagatcccattacccgtgtctataaccttgcgaagtctccatccccgttcgtatgcgcc  
actgttcacccgtqaacactacgcggctttgtcgagaagagggtactctgggttccgtatggcggtacaaaaccgtcgtaa  
atggcaggcggtatcaaagcggtagacggtaaatgatcacagttacccgtcaaggtaaagatgaagtgtcgctgact  
aatatccagaaggcgaaacctgggtccccactttaa (SEQ ID NO: 26)

12) ygbP  
atggcaaccactatttgatgttgcgccgtggttccggcgccggatttggccgtcgaatgc当地acggaatgtccataagcaa  
tatctctcaatcggtaaatcaaaccattctgaaacactcgggtcatgcgtgtggccatccccgggtgaaacgtgtcgatattgc  
cataagtccctggcgatagccgtttgcacaacttcctggcgaaatcatccgcaatcaccgttagatggccgtgatgagcgt  
gccgattccgtctggcagggtctgaaagccgtggcgacgcgcagtggttattgtgtcatgcacccgtcgcccttggcatc  
aggatgaccctcgccgatgtggcgtagcggaaaccagccgcacggggggatcctcgccgaccagtgcgcgatacta  
tgaaaacgtccgaaaccgggcaaaatgccaatgcataccgttagatgc当地acggatgtggccgtgacccgc当地aa  
ttccctcgtagctgttaca|gactgtctgacgcgcgtctaaatgaaggcgcgactattaccgc当地acggccgtcgatgg  
tattggcgatccatctcgatgtggcgtagcggataacatggactacgcgc当地ggaaatgtggccactggccgat  
gtttacccatcccgaaaccatccatcaggagaatacatata (SEQ ID NO: 27)

13) ybeY  
atgagtcagggtatccctcgatttacaactggcatgtaaagataattccgggttaccggaaagagagccagttcagacatggctg  
aatgcgggtatcccgcatgttccagaagaatcggaagtgacgatccgcgttgtcgataccggccaaagccacagtctgaatc  
tgacccatcgccgttaaggataagccgaccaactgtctcccccgttgaagtgcgcctggcatggaaatgcgtactgg  
gcgatctggttatctgcgcgtcagggttgtgagaaggaaagctcaggagcaaggcaaaccactggaggcgcactggcgcata  
tgggtgtgcacggcagtcgcatttgttaggttacgatcacatcgaaatgcgcagaaggaaaatggaaagccctcgaaac  
agagattatgcgtctgggtatgaggatccgtacattgcgcagaaggaaaataa (SEQ ID NO: 28)

15) kdtB  
atgaaaaacgggcgatttatccgggtactttcgatcccattaccalggtcataatcgatatcgacgcgcacgcagatg  
tcgatcacgttattctggcgattgccgcaccccgataaaaaaaaaccatgtttaccctgaaagagcgtgtggcaactggcacag  
caggcaaccgcgcacatgggaacgtggaaagtggcgggtagtgtatggcaacttcgccccgtaatcaacacgcac  
qqtqctgatcatggctacatcacatggcagatggaaatgtcaactggccatataatcccaacttaatggccggaa

ctggaaagtgttctgtgcgcgtaaaagagtggcgttatccttcatcggtggtaaaagagggtggcgccatcaggcg  
atgtcaccatccgtccggagaatgtccatcaggcgctgtatqcgaaqttaqcqtag (SEQ ID NO: 30)

16) pfs

17) ycaJ

18) b1808

Fig. 1 continued

tgtcgttgcgcgggtggcaccgcgttcgtatgaagtgcactaccaggatgcgcgtcattactctcaaacagggggttagggcgact  
gattcgcgacgcgcgacgatcggtggcggtgggtttgtacaatcggtggatgcgtcctacggcgacgttctcc  
agtctgcgcgcgcgcacgcacccgtgacattgcccgtcggttcctgtcgataccatccctccaggtaa (SEQ ID  
NO: 33)

19) yeaA

atggctaataacccttcggcagaagaactgaaaaaaaaattgtccgagatgcgcgtttacgtgacgcagaatcatgggacag  
aaccgcatttacgggtgtttactgcataacaaggcgacggcgatatactgttgtacgtcgatgcccccggttcatccc  
aaaccaagtatgttccggctgtggctggcccgatgttctacgaaccggtaagtgaagaatccatcgttatataaagactgtc  
acatggaatgcagcgatagaaattcgttgcgttaactgtatgcctatctgggcattgtctcccccgcggcccgagccaa  
cgggcgaacgttattgtgttaactctgcctttacgcgttaccgatggcggaaaacggcgaagaaatcaacggttga (SEQ  
ID NO: 34)

20) yagF

21) b1983

gtgggacgtaaatgggccaatattgtgtaaaaaaaaacggctaaagacggtgcaacgtctaaaattatgc当地  
gaaatctatgtctgtctaaacaaggtaaccccgatccagaattaaacacatctttaaaattcgatttgaacgtgcaagcag  
gcacaagtccaaagcacgttattgataaaagcaattgataaaagccaaaggccggagatgaaacggtcgtgcagggacgt  
tatgaaggcttggctcaatggctcaatgattatcgccgagacattgacttcaaatgttaaccgtacgatgtctaaacgttcgcaca  
attttcaataaaaaaggccgaatatggagccggcagggtctgtcagctatgttgcataatcgggtgtattttaaagg  
gacagaccctgaccatattttgaaattttacttgaagctgaaagtgtatgttcgtgttgtactgaagaagaaggtaacattttat  
tatactgaacctactgaccctcataaaggaaatcgccgtctaaaaggcagctggaatctactgagttctcaacaacagaatttaga  
aatgtatgtctcaatgtcaagttgagttccccagaagatgttagaaatcttgc当地  
atgtacaaaaaggatgtatcataacqtcqcaaattctctaa (SEQ ID NO: 36)

22) vidD

atggcgcggccactgtcgctggctcgccgggtctatgcgtatccctatcggttctatcaacgcgttattgtccgtactcgccgc  
cgccatgtcgtttactccaaacctgttcacgactacqqaatqaaqcattqccqcaqqtqgactqataaaaqccqagtgtttgac

6 / 15

gtgaaacgcgtataaaatgccacccttacaccctgggtggacgatcccgtccgccccgaccatttgataccagagaaca  
ctaa (SEQ ID NO: 37)

23) yceG

atgaaaaaaagtgttattgataatctgttattgcgtgtactgggtatgcgcgtgggtggcgtctggaaagggtcgccatctgc  
ccgacagcaaattgtttatcaaagaagagacgatatttaccctgaagccaggacggacgtctggcgctcggtaacagc  
ttatgcgataagatcatcaatgtccacgggtttcaatggctgtcgatcgaaaccggatcttcactttaaagccggac  
ttaccgttacaccgcagatgaccgtgcgcagatgtgaaattgtggaaagcggtaaagaagcacagtccctgtgc  
tggtagaaggatgcgtcgacgttacctaagaatgcgtgaggccccgtatatacgacatacgctgagcgatgataag  
tacgcacccgtacgcaggcactgtgaaaaccggagtggtggatgaaagggtggatctggccagacacctggatgtata  
ccgccaataccaccgcgtgcgttactcaagcgacgcacaagaaaatggtaaagcggtcgatagccctggaaagg  
gcgtgcgcacggctgcattaaagataaaaaaccgttggtgacgatggcatattcgaaaaagaaaccgcgttgc  
agtgaacgcataagggtgcctcgttattatcaaccgttacgcgttgcgcctgcagaccgaccgcaccgtgattacg  
ggatggagagcgttataatggcaaaacttctcgccgtcagaccgtggaaacgcgcacagcgatataaccctataccattaccgtt  
ctgcgcagggtgcgtacgcacgcggggcgattcgcgtgaaggctgtgcgcattccggaaaaacgcgttatcttatt  
ttgtggccgtatggtaaagggtgtcacacgttaataccatctgcgtcataacaagtgtgcaggattatcgaaagtgc  
aaggaaaaaaatgcgcagtaa (SEQ ID NO:38)

24) yjbC

atgcgtcccgactcatcagtccgtttaaataaaatacatcagcgaaagcggaaattgtcacgcgcgaagcggtatcgatata  
gagcaaggcaatgttcccttaatggcaagcgagccaccattggcgtcaggtaaaaccggcgacgttgtgaaagtaaac  
ggtcagttgtatgaaacctcggttgcggaaagccgaagatttgtactttatgcgcgttgcgcatacaagccgttggatattgt  
aaggatggagagcgttataatggcaaaacttctcgccgtcagaccgtggaaacgcgcacagcgatataaccctataccattaccgtt  
ctgcgcagggtgcgtacgcacgcggggcgattcgcgtgaaggctgtgcgcattccggaaaaacgcgttatcttatt  
ttgtggccgtatggtaaagggtgtcacacgttaataccatctgcgtcataacaagtgtgcaggattatcgaaagtgc  
aaggaaaaaaatgcgcagtaa (SEQ ID NO:39)

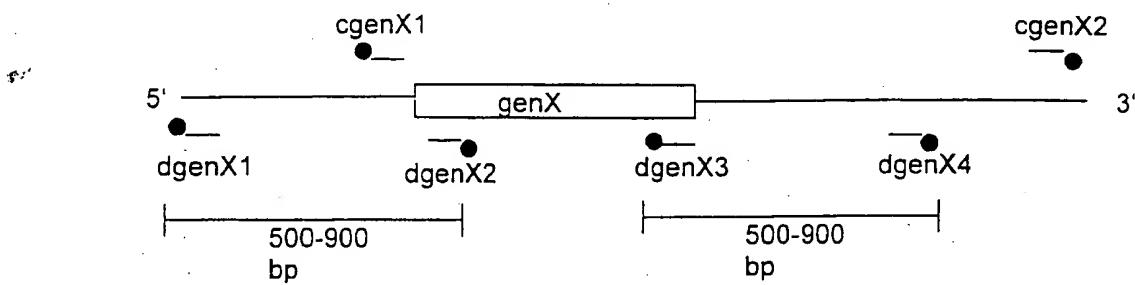


Fig. 2

E. coli	<i>B. subtilis</i>				<i>H. influenzae</i>				<i>H. pylori</i>			
	GeneRank# <sup>1</sup>	Gene name	SWIPI# <sup>2</sup> or Substitution# <sup>3</sup>	E-value	GeneRank# <sup>1</sup>	Score	E-value	GeneRank# <sup>1</sup>	Score	E-value	GeneRank# <sup>1</sup>	E-value
YgbB	g1789103	YacN	2Q06756	169	2,00E-43	1573672	205	1,00E-54	2314164	105	1,00E-24	
YhcC	g1789011	YaaJ	2P21335	135	2,00E-33	1573925	175	1,00E-45	2313814	24	3	
YacfE	g1786292	YtaG	2BG13824	135	2,00E-33	1573909	191	2,00E-50	2313965	87	4,00E-19	
YehB	g1787459	YahfA	2P37550	102	3,00E-23	1574450	317	3,00E-88	2314615	41	5,00E-06	
YefD	g1789510	YifF	2BG13940	126	6,00E-32	1574175	276	2,00E-74	2314637	61	5,00E-11	
Yih	g1789004	YieC	2P37565	71	1,00E-13	1573822	286	5,00E-79	2314107	24	5,4	
YggJ	g1789315	YjetJ	2P54461	96	2,00E-21	1573272	312	1,00E-86	2313478	49	1,00E-07	
YieL	g1790610	YdiB	2BG12199	89	3,00E-19	1573014	171	1,00E-44	2313840	46	9,00E-07	
YiaO	g1790004	YihfE	2BG12072	148	7,00E-37	1574060	374	1,00E-105	2314105	26	1,5	
YidC	g236210	YwlC	2P39153	93	3,00E-19	1573655	206	3,00E-56	2313122	29	0,13	
YihbC	g1789461	YksS	2P32726	90	1,00E-19	1574740	157	2,00E-40	2314193	65	1,00E-12	
YihbP	g1789404	YacM	2Q06755	129	3,00E-31	1573673	233	4,00E-63	2314164	47	5,00E-07	
YneY	g1786400	YqfG	2P46347	70	0,00E-14	1572948	190	2,00E-50	2314318	54	3,00E-06	
CysP	g1788653	YqfY	2P54482	318	4,00E-88	1573337	604	1,00E-174	2313753	294	3,00E-01	
KcdB	g1790065	Yihb	2BG13361	145	2,00E-36	1573650	176	5,00E-46	2314651	170	3,00E-44	
pls	g1786554	YifU	2BG13800	244	6,00E-66	1574146	254	2,00E-69	2313168	123	4,00E-30	
YeaJ	g1787119	YrvN	2BG13808	275	5,00E-75	1574435	660	0	2314168	201	3,00E-53	
b1008	g1788110	DinG	2P54394	245	6,00E-66	1573357	768	0	2313340	33	0,025,	
yeaV	g1788077	YppQ	2P54155	136	2,00E-32	1574293	116	3,00E-28	2313314	425	6,00E-31	
yafE	g1788654	lvd	2P51785	180	4,00E-46	1573744	168	7,00E-43	2314249	103	2,00E-23	
b1903	g1788294	YeeI	2BG12824	279	1,00E-76	1573285	161	2,00E-41	2313249	108	1,00E-25	
YidD	g140861	YjA	2BG13065	90	6,00E-19	1176311	96	1,00E-20	2314625	40	8,00E-04	

<http://www.ncbi.nlm.nih.gov/Entrez/protein.html><http://www.ncbi.nlm.nih.gov/ClustalW/clustalw.html><http://www.ncbi.nlm.nih.gov/Protein/ProteinList.html><http://www.ncbi.nlm.nih.gov/BLAST/finishedgenome.html>

Fig. 3

<i>E. coli</i>	<i>M. Infectiosiss</i>			<i>C. trachomatis</i>			<i>H. pylori</i>		
	Gene name	GenBank#	GeneRank#	score	Gene name	GenBank#	score	Gene name	GenBank#
Yqbl	g1789103	1077312	78	1,00E-16	3328865	62	1,00E-12	2688040	23
Yqbl	g1788911	2960176	119	1,00E-28	3329316	120	2,00E-29	2687969	23
Yqbl	g1786292	2113915	119	1,00E-28	3328928	62	1,00E-11	2688463	41
Yqbl	g1787459	2092140	84	1,00E-17	3329270	82	2,00E-17	2688545	23
Yqbl	g1788510	2336754	109	2,00E-25	3329180	105	0,00E-25	2688006	90
Yqbl	g1789004	1550650	25	6,40E+00	3329168	23	7,80E+00	2688577	23
Yqbl	g1789315	2078027	70	2,00E-13	3328922	23	8,60E+00	2688252	44
Yqbl	g1790610	1469365	60	1,00E-10	3328975	61	2,00E-11	2688077	60
Yqbl	g1790004	2113942	27	1,80E+00	3328868	26	1,30E+00	2688570	25
Yqbl	g2367240	1322425	68	0,00E-13	3328537	68	2,00E-13	2688669	73
Yqbl	g1789561	2078017	56	2,00E-09	3328787	26	4,00E-01	2688749	36
Yqbl	g1789104	1877313	99	3,00E-22	3328896	95	2,00E-21	2688781	26
Yqbl	g1786880	2078032	62	3,00E-11	3328852	26	4,20E-01	2687941	55
Yqbl	g1788663	2612013	277	9,00E-76	3328450	155	2,00E-39	2688019	31
KdhB	g1790065	1694866	140	8,00E-35	3329163	25	9,40E-01	2688628	97
pls	g1786354	1405762	100	1,00E-22	3328055	22	6,80E-01	2688288	152
YcaJ	g1787119	1460081	329	3,00E-91	3328753	60	1,00E-10	2688379	55
b1808	g1788110	1340095	274	1,00E-74	3329029	27	9,30E-01	2688551	30
YaaA	g1788077	1550715	126	8,00E-31	3328854	27	2,20E-01	2688358	23
YagF	g1786464	2213526	195	1,00E-50	3329033	28	7,30E-01	2688576	25
b1983	g1788294	2281051	124	9,00E-30	3328890	116	1,00E-28	2687898	138
Yqbl	g140861	28008707	73	1,00E-13	3328908	56	2,00E-08	2688025	52

Fig. 3 continued

<i>E. coli</i> gene name	GenBank# <sup>1</sup>	<i>T. pallidum</i>			<i>S. pneumoniae</i> *			<i>S. aureus</i> *		
		GenBank# <sup>1</sup>	score	E-value	contig# <sup>4</sup>	score	E-value	contig# <sup>4</sup>	score	E-value
Yqfb	g1789103	3322804	98	1,00E-22	/	n.d.	n.d.	/	n.d.	n.d.
YhhC	g1788911	3322548	33	4,00E-03	101	71	2,00E-12	49	102	2,00E-25
YiaeF	g1786292	3322572	36	6,00E-04	17	109	3,00E-24	/	n.d.	n.d.
YehB	g1787459	3322649	83	7,00E-18	/	n.d.	n.d.	/	n.d.	n.d.
YejD	g1788510	3322747	97	3,00E-22	41	166	2,00E-41	12	152	5,00E-37
YiiJ	g1789004	/	n.d.	n.d.	7	89	2,00E-15	249	82	9,00E-16
YggJ	g1789315	3322550	27	4,90E-01	93	65	1,00E-10	90	86	7,00E-17
YieI	g1790610	33223187	76	6,00E-16	140	80	2,00E-15	24	75	9,00E-14
YiaO	g1790004	3322480	28	2,90E-01	/	n.d.	n.d.	/	n.d.	n.d.
YidC	g2367210	3322447	39	6,00E-05	123	62	9,00E-10	193	76	3,00E-14
YihC	g1789561	3322709	26	4,40E-01	47	55	8,00E-08	173	90	1,00E-18
YihP	g1789104	3322804	58	2,00E-10	72	55	2,00E-07	/	n.d.	n.d.
YhcY	g1788800	3322948	48	1,00E-07	17	60	2,00E-09	396	75	7,00E-14
GcpE	g1788063	3322731	217	3,00E-58	/	n.d.	n.d.	/	n.d.	n.d.
KubB	g1790065	3322553	100	2,00E-23	232	113	2,00E-25	205	149	2,00E-36
pIs	g1790654	3322437	112	9,00E-27	156	102	5,00E-46	1235	82	1,00E-15
YcaJ	g1788719	3323329	53	1,00E-08	62	95	2,00E-23	1085	159	1,00E-30
b190B	g1788110	3322379	29	3,50E-01	114	114	5,00E-25	434	34	7,40E-01
yeaA	g1788077	3322932	114	1,00E-26	31	136	2,00E-33	422	112	1,00E-25
yauK	g17886464	3322975	26	3,20E-00	38	202	1,00E-51	24	171	4,00E-42
b1903	g1788294	3322762	142	9,00E-36	143	360	1,00E-99	412	183	2,00E-46
YellD	g140861	Treponema F71	4,00E-13	12	64	7,00E-11	1341	76	1,00E-14	

Fig. 3 continued

<i>E. coli</i> gene name	Genebank# <sup>1</sup>	<i>E. faecalis</i> *			<i>P. aeruginosa</i> *			<i>B. pertussis</i>		
		conseq# <sup>4</sup>	score	E-value	conseq# <sup>4</sup>	score	E-value	conseq# <sup>4</sup>	score	E-value
Yqbb	g1789103	6177	141	8,00E-34	93	181	5,00E-46	126	139	3,00E-33
YhAC	g1704911	6349	132	3,00E-31	93	151	7,00E-37	737	151	9,00E-37
YacE	g1786292	6196	111	1,00E-24	95	187	1,00E-47	924	159	3,00E-39
YcaB	g1787459	6342	114	2,00E-25	95	286	2,00E-77	1062	245	9,00E-56
YejD	g1788510	6176	137	2,00E-32	94	198	8,00E-51	983	91	1,00E-18
YrfH	g1789004	6199	97	2,00E-20	97	192	4,00E-49	1085	160	2,00E-39
Yq9J	g1789315	6207	75	1,00E-13	66	196	4,00E-50	551	119	4,00E-27
Yiel	g1790610	6294	29	4,00E-00	97	177	7,00E-45	762	125	4,00E-29
YiaO	g1790004	6236	125	1,00E-28	91	139	8,00E-33	459	201	1,00E-51
YiaC	g2367210	6208	96	4,00E-20	75	163	2,00E-40	362	43	4,00E-05
YihC	g1789561	6465	103	2,00E-22	85	148	6,00E-36	371	76	4,00E-14
YqlpP	g1789104	6311	55	2,00E-07	93	180	2,00E-45	126	93	5,00E-19
YoeY	g1786380	6206	67	1,00E-11	91	142	3,00E-34	369	89	5,00E-18
GcpE	g1789863	/	n.d.	n.d.	91	514	1,00E-145	862	161	2,00E-39
KlfB	g1790065	6304	147	1,00E-35	84	197	1,00E-50	1097	172	2,00E-43
pfs	g1786354	6395	201	1,00E-51	/	n.d.	/	n.d.	n.d.	n.d.
YcaJ	g1787119	6207	138	2,00E-32	89	529	1,00E-450	1043	452	1,00E-127
b1408	g1788110	6265	120	7,00E-27	82	215	1,00E-55	781	255	1,00E-67
ycaA	g1788077	6315	138	3,00E-33	81	158	2,00E-39	777	146	1,00E-35
yagf	g1786464	/	n.d.	n.d.	84	169	1,00E-41	759	160	8,00E-39
b1493	g1786294	6169	309	3,00E-04	82	145	5,00E-35	1059	155	6,00E-38
YidD	g140061	/	n.d.	n.d.	46	76	1,00E-14	1007	74	7,00E-14

Fig. 3 continued

<i>E. coli</i> gene name	GenBank# <sup>1</sup>	<i>B. subtilis</i> GenBank# <sup>1</sup>	score	E-value	<i>H. influenzae</i> GenBank# <sup>1</sup>	score	E-value	<i>H. pylori</i> GenBank# <sup>1</sup>	score	E-value
yceG	91787339	92635201	140	2e-32	91073838	289	2e-77	gbAAD07652.1	87	3e-16
yhcC	9396357	92634751	132	3e-30	91574128	101	7e-21	g2314637	99	4e-20

<i>E. coli</i> gene name	GenBank# <sup>1</sup>	<i>M. tuberculosis</i> GenBank# <sup>1</sup>	score	E-value	<i>Ch. trachomatis</i> GenBank# <sup>1</sup>	score	E-value	<i>B. burgdorferi</i> GenBank# <sup>1</sup>	score	E-value
yceG	91787339	embCAB06185	74	2e-12	/	n.d.	n.d.	92688649	101	1e-20
yhcC	9396357	g2326754	110	1e-23	g3329180	132	2e-30	g2688006	91	1e-17

<i>E. coli</i> gene name	GenBank# <sup>1</sup>	<i>T. pallidum</i> GenBank# <sup>1</sup>	score	E-value	<i>S. pneumoniae</i> <sup>*</sup> config# <sup>4</sup>	score	E-value	<i>S. aureus</i> <sup>*</sup> config# <sup>4</sup>	score	E-value
yceG	91787339	g3322780	108	7e-23	/	n.d.	n.d.	/	n.d.	n.d.
yhcC	9396357	g3322747	90	2e-17	12	141	5e-34	4402	134	5e-32

<i>E. coli</i> gene name	GenBank# <sup>1</sup>	<i>E. faecalis</i> <sup>*</sup>	score	E-value	<i>P. aeruginosa</i> <sup>*</sup> config# <sup>4</sup>	score	E-value	<i>B. pertussis</i> <sup>*</sup> config# <sup>4</sup>	score	E-value
yceG	91787339	6216	115	4e-26	54	231	7e-61	398	209	2e-54
yhcC	9396357	6178	118	4e-27	54	109	2e-24	190	123	9e-29

Fig: 3 continued

Multiple sequence alignment of E. coli gene ygbB with 5 orthologs from different organisms  
 Legend: 1 = Escherichia coli; 2 = Haemophilus influenzae;  
 3 = Bacillus subtilis; 4 = Synechocystis; 5 = Treponema pallidum;  
 6 = Helicobacter pylori; 7 = Alignment score (\* identical :  
 chemically similar . sterically similar)

1	-
2	-
3	-
4	-
5	MRRGGACVQKKEYLPLTSRQPVGVCLLSEILVRALEARSFFFLVVVVTVPAGEVAYAESQVAC
6	-MSLIRVNGEAFKLSSLSEEDPFETKETLETLIKQTSVVLLAAGESRRRFSTTIKKQWLR
7	
1	-
2	-
3	-
4	-
5	DSRLSAFPQRSTRPVILYVPGGAHTRSASVRAGLDAMATHAPDVVLVHDGARPFSVALIHS
6	SNHTPLWLSSVYESKEALDFKEIILVVSELDYIYIKRHYPEIKLVKGASRQEHSVNAIK
7	

Fig.

4

Fig. 4 continued

1 VLEATCRYGAAPVILEATDTPKGVAADGSIETHLIRSRVRLAQTPOQFCYASLCAAHHRA  
2 -  
3 -  
4 -  
5 TIDSAYTLLTSDVARGLANIEALKNLFLTLQQTSHYCTIAPYLPYCYDTAIYYNEALDREA  
6 IK -  
7

1 -  
2 -  
3 -  
4 -  
5 ATDGEOQYTDDSELYARYG---GTUVHVCAGERSNVKITYPEPDLERQRASEPALTRGIVS  
6 LIQTPQLSHTKALQSALNQGDFEKDESSAILQAFPDRVSYIEGSKDLHKLTTSGDLKHFTL  
7

1 ---MРИGHGFDVHAFGGEGPIIIGVRIPIYEKGILLAHSDDGVALTHALTDA  
2 ---MIRIGHGFDVHAFGFEDRPLIIGVVEVPYHTGFIAHSDGDVALHALTD  
3 ---MFRIGQGFDVHQLVEGRPLIIGGIEIPYEKGILLGHSDDVLLHTVADA  
4 ---MTALRIGNGYDIHRLVGDRPLIILGGVTIAHHLGIDGHSDDVLT  
5 PCTEEGALRVGLGTDMHALCAGRPLIILAGIHIPSKKGAQGHSDDVLA  
6 FENPAKDTFIGMGFDTTHAEIKDKPMVLGGVVLDCEFGLKAHSDDGDALL  
7

Fig. 4 continued

1	LGDIGKLF	PDT	DPAFKGADSRELLREAWRR	IQAKGYTLGNVDVTIIAQAPKMLPHIPQMR
2	LGDIGKLF	PDT	DMQYKNADSRG	LLREAFRQVQEKGYKIGNVDITIIAQAPKMRPHIDAMR
3	EGDIGKH	FPTD	DPEF	KDAFSFKLLOHVGIVKQKGVLGNIDCTIIAQPKMLPYIEDMR
4	LGDIGHYEF	PPSDARWQGADSL	KILLAQVHQLL	ERGWIRINNLDNVVAEQPQLKPHIQAMIK
5	LGDIGT	EEFPSCDG	RWKDAHSCALLRHTWQLVR	VRAACWRLVNLDAVVCLEQPALHPFREAMR
6	GGDIGEWFPDNDP	DKRYKNASSSKELLKIVLDFSQSIGFFELFEMGATI	SEIPKITYKPATI	
7	* * * *	* : : . * * *	: : : : : : : : : : : :	: : : : : : : : : : : :
1	VFI	IAEDLGCHMDD	DVNVKATTTEKL	GFTGRGEGIACEAVALLIKATK-
2	AKIAEDL	QCDIEQVNVKATTTEKL	GFTGRQE	GIACEAVALLIRQ---
3	KRIAEGL	ADVSQVNVKATTTEKL	GFTGRAEGIAA	QAQATVLIQKG---
4	ENLAKVLT	IDPDLLIGIKATTNERL	GPTGREEGIAAYSV	ALLIKEG---
5	ASLAQALDTH	VTRVFVKAKTAERL	GSGAAVTAQVVVLLKKI	--
6	ENLSQLLGLEKSQ	ISLKATTMEKM	GFIGKQEGLLVQAHVSMRYKQKL	
7	:: : *	. : : * . * : : * *	(	: : : : : : : : : : : :

1

## SEQUENCE LISTING

&lt;110&gt; GPC Biotech AG

&lt;120&gt; Novel method for identifying antibacterial compounds

&lt;130&gt; D 1400 PCT

&lt;140&gt;

&lt;141&gt;

&lt;160&gt; 45

&lt;170&gt; PatentIn Ver. 2.1

&lt;210&gt; 1

&lt;211&gt; 39

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: artificial sequence

&lt;400&gt; 1

atggttataaa atttggagtg tgaaggttat tgcgtgtaa

39

&lt;210&gt; 2

&lt;211&gt; 33

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: artificial sequence

&lt;400&gt; 2

gttataaaatt tggagtgtga aggttattgc gtg

33

&lt;210&gt; 3

&lt;211&gt; 33

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: artificial sequence

&lt;400&gt; 3

cacgcaataaa ccttcacact ccaaatttat aac

33

&lt;210&gt; 4

&lt;211&gt; 36

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

2

<220>  
<223> Description of Artificial Sequence: artificial  
sequence

<400> 4  
gatcggatcc aaattccagt tagccatgat gcggtc

36

<210> 5  
<211> 54  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial  
sequence

<400> 5  
cacgcaataa cttcacact ccaaatttat aaccattata cacggacgct atgc

54

<210> 6  
<211> 55  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial  
sequence

<400> 6  
gttataaaatt tggagtgtga aggttattgc gtgacggatt aattttgttt ctctt

55

<210> 7  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial  
sequence

<400> 7  
gatcgctcgac gcgctcgata tcacccatga acaaccg

37

<210> 8  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial  
sequence

<400> 8  
caatccgctg ctttatttct gtcag

25

<210> 9  
<211> 25  
<212> DNA  
<213> Artificial Sequence  
  
<220>  
<223> Description of Artificial Sequence: artificial sequence

<400> 9  
ttataacgaa atcaacggga aacct

25

<210> 10  
<211> 34  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial sequence

<400> 10  
gatcggatcc ctctaaaaaa taaggaatta aagg

34

<210> 11  
<211> 56  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial sequence

<400> 11  
cacgcataa cttcacact ccaaatttat aaccatagga tacctaatta attaac

56

<210> 12  
<211> 54  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: artificial sequence

<400> 12  
gttataaatt tggagtgtga aggttattgc gtgaagagcg ccatttccca ccgt

54

<210> 13  
<211> 36  
<212> DNA  
<213> Artificial Sequence

<220>

4

<223> Description of Artificial Sequence: artificial sequence

<400> 13  
gatcgctcgac tcataattgtc gataacccgc tgcggt

36

<210> 14

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: artificial sequence

<400> 14

gttgacggcc aggccaacag tcat

24

<210> 15

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: artificial sequence

<400> 15

ataaccctgg gcaatcgcc tgg

22

<210> 16

<211> 480

<212> DNA

<213> Escherichia coli

<400> 16

atgcgaattt gacacggttt tgacgtacat gcctttggcg gtgaaggccc aattatcatt 60  
ggggcgtagc gcatttccta cggaaaaggaa ttgctggcg attctgatgg cgacgtggcg 120  
ctccatgcgt tgaccgatgc attgcttggc gccccggcg tggggatat cggcaagctg 180  
ttccccggata ccgatccggc atttaaagggt gccgatagcc gcgagctgct acgcgaagcc 240  
tggcgtcgta ttcaggcgaa gggttataacc ctggcaacg tcgatgtcac tatcatcgct 300  
caggcaccga agatgttgcc gcacattcca caaatgcgcg tggatttgc cgaagatctc 360  
ggctgccata tggatgtatgt taacgtgaaa gccactacta cggaaaaact gggatttacc 420  
ggacgtgggg aaggattgc ctgtgaagcg gtggcgctac tcattaaggc aacaaaatga 480

<210> 17

<211> 537

<212> DNA

<213> Escherichia coli

<400> 17

atgcggccgc cttttataac cggagtttc tttttgtctg aagtcaatt tagccacgaa 60  
tactggatgc gtcacgcgt gacgctggcg aaacgtgcct gggatgagcg ggaagtgcgg 120  
gtccggcgcc tattaaatgcg taacaatcg gtaatcgcc aaggctggaa ccccccgtt 180  
ggtcgcctatg atccaccgc acatgcagaa atcatggccc tgcggcaggg tggctctgg 240

5

atgcaaaatt atcgctctgat cgacgccacg ttgtatgtca cgcttgaacc atgtgtaatg 300  
 tggccggag cgatgatcca cagtcgcatt ggtcgcgtgg tctttgggtc gcgtgacgcg 360  
 aaaactggcg ctgcgggatc ttaatggat gtgctgcata atccgggtat gaatcaccga 420  
 gtggaaatta cggaaaggaat actggcgat gagtgcgcgg cggtgcgtcag tgacttcttt 480  
 cgcacatgcgcg gccaggaaat taagcgcag aaaaaagcgc aatcctcgac ggattaa 537

<210> 18  
 <211> 621  
 <212> DNA  
 <213> Escherichia coli

<400> 18  
 atgaggata tagttgcctt aacgggaggc attggcagt gcaagagtac cgttgccaat 60  
 gcgtttctg atctcgaaat taacgtcatt gatgccata ttattgcgcg tcaggtggtt 120  
 gaaccagggtg cacctcgct acatgccatt gctgatcact ttggcgctaa catgattgct 180  
 gctgatggaa cattgcagcg cccggccttgc cgcgagcggta tcttcgc当地 cccggaagag 240  
 aaaaactggc ttaacgcctt gtcgtatccg ctgattcagc aagagacgca acaccagatc 300  
 cagcaagcta ctccccctt tgcgtatgtgg gttgtgc当地 tgctgttaga aaactcactg 360  
 tataaaaaag cgaatcgatg gttgtgtggg gatgtcagcc cagaaacgca acttaagcgc 420  
 accatgcagc gcgtatgtgt aactcgcgag catgtcgaac aaatccttgc tgctcaggca 480  
 acgcgcgaag cccgccttgc cgtggcagat gacgtcattt ataataacgg cgcacccggat 540  
 gctatcgat cggatgttgc cccgcctgcac gcacactatt tgctcaggca gtcgcagttt 600  
 gtctcacagg aaaaacctgt a 621

<210> 19  
 <211> 852  
 <212> DNA  
 <213> Escherichia coli

<400> 19  
 atgcggacac agtggccctc tccggcaaaa cttaatctgt ttttatacat taccggtcag 60  
 cgtgcggatg gttaccacac gctgcaaacg ctgtttcaat ttcttgatta cggcgcacacc 120  
 atcagcattt agtttcgtga cgatggggat attcgtctgt taaccccgt tgaaggcgtg 180  
 gaacatgaag ataacatgtat ctttcgcgc ggcgcgtatgt tgataaaaac tgccgc当地 240  
 agcgggcgtc ttccgacggg aagcgggtgcg aatatcagca ttgacaagcg tttgccc当地 300  
 ggcggcgtc tccgggtgg ttcatccat gccgc当地 cggg tcctgggtgc attaaatcat 360  
 ctctggcaat gcgggctaaag catggatgag ctggcggaaa tgggctgac gctggcgc当地 420  
 gatgttctg tctttttcg ggggcatgcc gcgtttccg aaggcgttgg taaaatacta 480  
 acgcgggtgg atccggccaga gaagtggat ctggtggc当地 accctgggtg aagtattccg 540  
 actccgggtga tttttaaaaatc tccgtactc cccgc当地 aata cgccaaaaag gtcaatagaa 600  
 acgttgtcaa aatgtgaatt cagcaatgtat tgcgaggta tcgcaagaaa acgttttgc当地 660  
 gaggttgtat cgggtcttgc ctggctgtta gaatacgc当地 cgtcgc当地 gactggaca 720  
 ggggcctgtg tctttgtga attgtatatac gagtctgaag cccgc当地 aggt gctagagcaa 780  
 gccccggaaat ggctcaatgg ctttgc当地 ggccgaaat ggctcaatgg ctttgc当地 ggcc 840  
 gccatgc当地 aa 852

<210> 20  
 <211> 696  
 <212> DNA  
 <213> Escherichia coli

<400> 20  
 atgcgacttg ataaatttat cgcacagcaa ctcggcgtta gccgtctat tgccggggcgt 60  
 gaaatccgc当地 gcaatcgatg caccgtcgat ggc当地 aatcg tccgtatgc agcgttcaaa 120  
 ctgc当地 cctg aacatgtatg cgttacgtat ggcaacccgc tggcgc当地 acacgggtcc 180  
 cgttacttca tgctcaatgg ccctcaggcc tatgtttgtat ccacggacgc ccctgatc 240

```
ccaacgggtgc tctatTTTct tgatgaacccg gtagcgtggaaactgcattgc ggccggggccgg 300  
ttggatattg ataccaccgg tctgggtctg atgactgtatggtcgttg gtcgcacccg 360  
attacttctc cgcccatca ttgcgagaag acctatctgg tgacactgttgcataacgaaaagatctc 420  
gctgacgata cggcagagca atttgctaaa ggcgtgcaggc tgcatatcgaaa 480  
actaagcctg cgggtgtggaa agtgattacc ccaacgcagg ttctgtctgac catcagcgaa 540  
gggcgttatac atcagggtgaa acgcattgttc gcccgggtgg gtaaccacgttggtagactg 600  
catcgtgaac gtattggccgg tattacgtgt gatgtctgatt tagccccccgg tgaatatcg 660  
ccgttaactg aagaagaaat tgccagcgctc gtctaa 696
```

<210> 21  
<211> 885  
<212> DNA  
<213> Escherichia coli

<400> 21  
atgattatgc cgcaacatga ccaattacat cgctatctgt ttggaaaactt tgccgtgcgc 60  
ggcgaactgg taaccgtttc ggaaaccctcg caacagatcc ttgagaacca cgattatccg 120  
cagccccgtta aaaacgtgct ggcagaactg ctgggttcgca ccagcctgtt aaccgctacg 180  
ctgaagtttg atggtgatata caccgtacag ctgcaggcg acggccgtt gaatctggc 240  
gttattaacg gtaacaataa ccagcagatg cgccgtgtgg cgccgtgcga gggcgaaatt 300  
ccagaaaatg ccgacacctgaa aacgctggtc ggcaatggtt acgttgttat caccattacc 360  
ccgagcgaag gcgaacgcta tcagggcgta gtgggtctgg aagggtgatac cctggcgccc 420  
tgcctgaaag attactttat gcgttctgaa cagctggca cgcgcctgtt tattegcac 480  
ggcgcacgtag acggcaaacc ggctgcaggc ggtatgtgt tgccaggataat gcctgcgc 540  
aatggcccaacg aggacgactt tgaccacccgt ggcacgttaa ccgaaaccat caaaaccggaa 600  
gaactgtgtga ccttaccggc aaaccaagtg ttgtggcggtt tgcgtatccgaa agaagagggt 660  
acgggtttagt atcccgaggaa ttgtggagttt aatgcacccgtt gcttcgtgtca acgttgcgc 720  
gatgcgtgtca aaacgctgcc ttgtgaagaa gttgatagca tcctggcggtt agatggcgaa 780  
attgacatgc attgtgattt ctgcgttac cactatctgt tcaatgcgtt ggatattgtt 840  
gaaatccgca acaacccgtc tccggcgat ccgcgttattaa 885

<210> 22  
<211> 759  
<212> DNA  
<213> Escherichia coli

```

<400> 22
gtggggagac gacgcggatt ttaactatg cgtatcccc gcatttatca tcctgaacca 60
ctgaccagcc atttcacat cgcgcttgc gaagatgccg ccaaccatat cgggcgcgtg 120
ctgcgcatgg ggccggggca ggcgttgcaaa ttgtttgacg gtagcaacca ggtctttgac 180
gccgaaatta ccagcgcag caaaaaaaaaaagc gtggaagtga aggtgtggc aggccagatc 240
gacgatcgcg aatctccgct gcatattcac ctcggctcagg tgatgtcgcg tggtgaaaaaa 300
atggaattta ctatccagaa atcgatcgaa ctcgggtgtaa gcctcattac gccactttt 360
tctgagcgct gccggcgtaa actggatagc gaacgctcta acaagaagat tcagcgttgg 420
cagaagatgg caatgtctgc ctgtggcgac tgggtgtgta accgggtgc gaaatccgt 480
ccacgcgtgg atctggaaagc ctgggtgtca gaggcaggatc aaggactgaa actgaatctt 540
cacccgcgcg ccagtaaacag catcaatacg ttgcggttac cgggtgaacg cgtccgcctt 600
ctgattggcc cggaaaggcggtt tttatcgca gatgaaaattt ccatgactgc ccgttatcaa 660
tttactgata tcctgttggg acctcgcggtt ttgcgtacag agacaactgc gtcaccgc 720
attaccgcgc tacaagtacg atttggcgat ttgggtctaa 759

```

<210> 23  
<211> 462  
<212> DNA  
<213> Escherichia coli

&lt;400&gt; 23

atgatgaatc gagtaattcc gctccctgat gagcaggcaa cattagaccc gggcgagcgg 60  
gtagcgaaag cctgcgatgg cgcacccgta atctatctgt atggcgattt aggcgcaggt 120  
aaaaccaccc tttagccccgg cttttacag gctctgggtc atcaggtaa tgtaaaaagc 180  
cccactata cgctggcga accctatacg ctgcacaact taatggtcta tcaactttgat 240  
ttgtaccgcc ttgccgatcc cgaggagctg gagtttatgg ggatccgca ttatccgccc 300  
aacgtgcga tctgccttgtt ggagtggcca caacaaggta cagggttct tcctgaccgg 360  
gatgtcgaaa tacacattga ttatcaggca caaggccgtg aggcgcgcgt gatgtcggtt 420  
tcctctgcgg gtgaatttgtt gctggcggtt tttagccggaa aa 462

&lt;210&gt; 24

&lt;211&gt; 987

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;400&gt; 24

atggaaattac gctctgttaac ctacgcatta ttcattgtcg gcctggctgc attcagcaca 60  
tcttctctgg cggcacaatc tttacgtttc ggttatggaa catcacaaac cgactcgcaa 120  
catattgcgg cgaaaaattt caatgattta ttgcaggaga gaaccaaagg cgagctgaaa 180  
ttaaaactgt tcccgacac cactctcggt aacgcgcagg cgatgatcg cggcgatcg 240  
ggcggcacca tcgatatggaa aatgtccggc tcgaataact ttgccccgtt attaccagg 300  
atgaacttgc ttgatgtccc tttccgttgc cgcgataccg ctcacgcgc 360  
gacggcaaaag tcgggtatga tctgaaagcc tcacttgaag gtaaaggact gaaagtactg 420  
gcctactggg aaaacggctg gcgcgatgtc accaactcgc ggcgcacccggtaaaaaccccc 480  
gccgacactga aagggtctgaa aatccgcacc aacaatagcc cgatgatata cgccgcattc 540  
aaagtctttg ggcgtaaccc gatcccgatcg cggcttgcgc aagtctatac cgggctggaa 600  
acccgcacta tcgacgctca ggaacaccccg atcaacgtcg ttcgtgcgcgaaaattttc 660  
gaagtgcaga agttctttc tctgacgcac cacgcctatt ccccgcttctt ggtggtgatc 720  
aacaaggcga agtttgcattt cttaaatccgcg gagggtccgc aggcgcctgtt ttcatctgca 780  
caagaaggcg gtaactatca ggcgcactg gttgctgaag atcagcaaaa aatcatcgac 840  
ggcatgaaag aaggccccgt ggaagtcatc accgatctcg accgcacccgc ctttagcgac 900  
gcactgggaa atcaggatcg cgacatgtttt gttaaatggatcg tccgcacccgg agctgatctg 960  
ctgaaagccg tggatgaggt gcaataaa 987

&lt;210&gt; 25

&lt;211&gt; 573

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;400&gt; 25

gtgaataata acctgcaaaag agacgcatac gcagctgcga tagatgttct caatgaagaa 60  
cggtgtcatcg cctatccaac ggaagccgtt ttcgggtttt ggtgcgatcc tgatagcgaa 120  
acagcagtga tgcgactgtt ggagttaaaa cagcgtccgg ttgataaggg gctgattttt 180  
atcgcgacaa attacgagca gcttaaaccct tatattgtatg acaccatgtt gactgacgtg 240  
cagcgtggaaa ccatttttcc cgctggcca ggtccctgtca cctttgtctt tcccgccct 300  
gcgacaacac cgccgtgggtt gacggggccgc ttgatttcgc ttgctgtacg agtcaccgc 360  
catccgttgg tgggtgtttt gtgcccaggct tatggtaac cgcgttgc taccagtgc 420  
aacttgatgtt gattgccacc ttgtcgaaca gtagacgaag ttcgcgcaca atttggcg 480  
gcgttcccg ttgtgcctgg tgaaacgggg gggcgttaa atccttcaga aatccgcgt 540  
gcctgcacgg gtgaactgtt tcgcacaggaa taa 573

&lt;210&gt; 26

&lt;211&gt; 459

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

8.

&lt;400&gt; 26

gtgggctgt ccacattaga gcaaaaatta acagagatga ttactgcgcc agttgaggcc 60  
 ctgggtttt aactggttgg catcgattt attcgcggtc gcacatccac actgcgcata 120  
 tatattgata gtgaagatgg catcaatgtt gatgattgtg ctgatgttag ccaccaggta 180  
 agtgctgtgc tggatgttga agatcccattt accgttgtt ataaccttga aytctcctca 240  
 cccggctctcg atcgcactt gttcacggc gaacactacg cccgtttgt cggagaagag 300  
 gtgactctgg ttctccgtat ggccgtacaa aaccgtcgta aatggcaggg cgttatcaaa 360  
 gcggtagacg gtgaatgtat cacagttaacc gtcgaaggtt aagatgaagt gttcgctg 420  
 agtaatatcc agaaggcgaa cctgggtcccc cacttttaa 459

&lt;210&gt; 27

&lt;211&gt; 711

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;400&gt; 27

atggcaacca ctcatttggta tgtttgcgcc gtggttccgg cggccggatt tggccgtcga 60  
 atgcaaacgg aatgtcctaa gcaatatctc tcaatcggtt atcaaaccat tcttgaacac 120  
 tcgggtcatg cgctgtctggc gcattccccgg gtgaaacgtg tcgtcattgc cataagtcc 180  
 ggcgatagcc gtttgcaca acttccctgt gcaaatcatc cgcaaatcac cgttgttagat 240  
 ggcgggtatg agcgtgcccga ttccgtctgt gcagggtctga aagccgctgg cgacgcgcag 300  
 tgggtattgg tgcgtgacgc cgctgtctt tgtttgcattc aggatgaccc cgcgcgattt 360  
 ttggcgttga gcaaaaccagg cccgacacggg gggatccctcg cccgaccaggat ggcgcataact 420  
 atgaaacgtg ccgaaccggg caaaaatgcc attgtctata ccgttgcattcg caacggctta 480  
 tggcacgcgc tgacgcgcga atttttccct cgtgagctgt tacatgtactg tctgacgcgc 540  
 gctctaaatg aaggcgcgac tattaccgac gaaggccttgc cgttgcatttgc 600  
 catcctcaatg tggcgttgcagg ccgtgcggat aacattaaatg tcacgcgcggc ggaagattt 660  
 gcaactggcc agtttacct caccgcggacc atccatcagg agaatacata a 711

&lt;210&gt; 28

&lt;211&gt; 468

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;400&gt; 28

atgagtcagg tgatcctcga tttacaactg gcatgttgcgg ataattccgg gttaccggaa 60  
 gagagccagt ttcagacatg gctgaatcg ggtatcccgc agtttcaggaa agaattcgaa 120  
 gtgacgatc gctgtgtcgta taccggccgaa agccacacgt tgaatctgc ctatcgccgt 180  
 aaggataacgc cgaccaacgt gctctccttc ccgttgcgg tgcgcgcctgg catggaaatq 240  
 tcgctactgg gcatctgtt tatctgcgtt caggtgttgc agaaggaaacgc tcaggagcaa 300  
 ggcaaaaccac tggaggcgca ctgggcgcattt atgggtgtgc acggcgttgc cattttgtta 360  
 gtttacgatc acatcgaaaga tgacgaagca gaagaaatgg aagccctcgaa aacagagattt 420  
 atgcttgcgtt tggcttatgaa ggatccgttac attgcccggat aagaataaa 468

&lt;210&gt; 29

&lt;211&gt; 1119

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;400&gt; 29

atgcataacc aggctccaaat tcaacgttgcgg aaatcaacac gtatttacgt tggaaatgtt 60  
 ccgattggcg atgggtgtcc catcgccgtt cagtccatgtt ccaatacgcc tacggacac 120  
 gtcgaagcaa cggtaatca aatcaaggcg ctggaaacgcg ttggcgctgtt atcgccgtt 180  
 gtatccgtac cgacgttgcgaa cgcggccggaa gcgttcaaaac tcatcaaaaca gcaggtaac 240  
 gtggccgttgcgg tggctgttgcattt ccacttcgac tatcgccattt cgttgcggaaatgg 300  
 ggcgttgcgtt ttctgttgcgtt taacccttgcggat aatatcggtt atgaagagcg tatttcgtt 360

gtggttgact gtgcgcgcga taaaaacatt ccgatccgta ttggcgtaa cgccggatcg 420  
ctggaaaaag atctgcaaga aaagtatggc gaaccgacgc cgcaggcggt gctggaatct 480  
gccatgcgtc atgttgcata tctcgatcg ctgaacttcg atcagttcaa agtcagcggt 540  
aaagcgctcg acgtcttcg cgtgttgag tcttatacgat tgctggcaaa acagatcgat 600  
cagccgttgc atctggggat caccgaagcc ggtggcgcc gcagcccccc agtaaaatcc 660  
gccattgggt tagtgtctgtc gctgtctgaa ggcatcgccg acacgctcg cgtatcgctg 720  
gcggccgatc cggtogaaga gatcaaagtc gggttcgata ttttggaaatc gctgcgtatc 780  
cggtcgcgag ggatcaactt catcgccgtc ccgacctgtt cgcgtcgagga atttgcgtt 840  
atcggtaggg ttaacgcgtt ggagcaacgc ctggaaagata tcatactcc gatggacgtt 900  
tcgattatcg gctgcgttgc gaatggccca ggtgaggcgcc tggttctac actcggcgatc 960  
accggcgccca acaagaaaaag cggccttat gaagatggcg tgcgcaaaaga ccgtctggac 1020  
aacaacgata tgatcgacca gctggaaagca cgcattcgat cgaaagccag tcagctggac 1080  
gaagcgcgtc gaattgacgt tcagcagggtt gaaaaataaa 1119

&lt;210&gt; 30

&lt;211&gt; 480

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;400&gt; 30

atgczaaaac gggcgattta tccgggtact ttcgatccca ttaccaatgg tcataatcgat 60  
atcgtgacgc gcgcacgcga gatgttgcgt cacgttatttc tggcgattgc cgccagcccc 120  
agtaaaaaac cgatgtttac cctggaaagag cgtgtggcac tggcacagca ggcaaccgcg 180  
catctggggat acgtggaaatg ggtcgggttt agtgattttaa tggcgaactt cgcccgtaat 240  
caacacgcta cggtgctgtat tcgtggcctg cgtgcgggtt cagattttga atatgaaatg 300  
cagctggcgatc atatgaatcg ccacttaatg ccggaaacttgg aaagtgtgtt tctgtgcgg 360  
tcgaaagaaatg ggtcggttat ctcttcatcg ttggtgaaag aggtggcgcc ccatcaggcc 420  
gatgtcaccatc atttcctgccc ggagaatgtc catcaggcgcc tgatggcgaa gtagcgtatc 480

&lt;210&gt; 31

&lt;211&gt; 699

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;400&gt; 31

atgaaaatcg gcatcattgg tcaatggaa gaagaagttt cgcgtctgc tgacaaaatc 60  
gaaaaccgtc aaactatcg tctcgccgtt tgcgaaatct ataccggcca actgaatgg 120  
accggagggtt cgcttctgaa atcgggcata ggttaaagtcg ctgcggcgct gggtgccact 180  
ttgctgttgg aacactgtc gccagatgtt attattaaca ccggttctgc cggcggcctg 240  
gcaccaacgt tgaaaatggg cgtatcgat tgcgtcgacg aagcaacgtt tcaacgacgcg 300  
gatgtcaccgg catttggta tgaatacggt cagttaccag gctgtccggc aggttttaaa 360  
gctgacgata aactgatcg tgcgcgttag gcctgcattt ccgaactgaa tcttaacgct 420  
gtacgtggcc tgattgttag cggcgcacgtt ttcatcaacg gttctgttgg tctggcgaaa 480  
atccggccaca acttcccaca gcccattgtt gtagagatgg aagcgcacggc aatcgcccat 540  
gtctgccaca atttcaacgt cccgtttgtt gtcgtacgc ccattccgaa cgtggccgat 600  
caacagtctc atcttagctt cgtatgatgtt ctggctgttgc cgcctaaaca gtccagcctg 660  
atggttgagt cactgggtca gaaacttgca catggctaa 699

&lt;210&gt; 32

&lt;211&gt; 1344

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;400&gt; 32

gtgagcaatc tgtcgtcgat ttttcggat aataacttttc aacctctggc cgcgcgtatg 60  
cgcccgaaaa atttagcaca gtatatcgat cagcaacatt tgctggctgc ggggaagccg 120

ttgccgcgcg ciatcgaagc cgggcattta cattctatga tcctctgggg gcccgggg 180  
 accggcaaaa caactctcg tgaagtgatt gcccgtatg cgaacgctga tgtggaacgt 240  
 atttctgccg tcacctctgg cgtaaagag attcgcgagg cgatcgagcg cgcccgaa 300  
 aaccgcaatg caggcgccg cactattct tttgttgcg aagttcacccg tttcaacaaa 360  
 agccagcagg atgcatttc gccacatatt gaagacggca ccatactt tattggcga 420  
 accactgaaa accgtcggt ttagcttaat tcggactgc tttccgtgc ccgtgtctat 480  
 ctgttgaat ccctgagttac agaggatatt gagcaagtac taactcaggc gatggaaagac 540  
 aaaaccctgt gctatggtgg tcaggatatt gtctgtccag atggaaacacg acgcgcatt 600  
 gctgaactgg tgaatggcga cgcgcgcgg gcgttaataa cgctggaaat gatggcgat 660  
 atggccgaag tcgatgtatag cggtaagcgg gtcctgaagc ctgaattact gaccgaaatc 720  
 gccggtgaac gtacgcggccg ctttgataac aaaggcgatc gctttacga tctgatttcc 780  
 gcactgcata agtcgttacg tggtagcgc cccgatgcgg cgctgtactg gtatgcgcg 840  
 attattacg ctgggtggcga tccgttatat gtcgcgcgtc gctgtctggc gattgcgtct 900  
 gaagacgtcg gtaatggcga tccacggcgc atgcaggatgg caattgcggc ctgggattgc 960  
 tttaactcgcg ttggccggc gaaagggtgaa cgcgcatttgc ctcaggcgat tggtaacctg 1020  
 gcctgcgcgc caaaaagcaa cgctgtctac actgcgttta aagccgcgtt ggcgcgtatgc 1080  
 cgcgaacgcc cggattatga cgtgcccgtt cattgcgtt atgcgcgcac gaaattaatg 1140  
 aaggaaatgg gctacggcga gaaatatcg tacgctcatg atgaagcaaa cgcttatgtc 1200  
 gccggtgagg tttaacttccc gccggaaata gcacaacac gctattatcc cccgacaac 1260  
 aggggccttg aaggcaagat tggcgaaaag ctcgccttgc tggcgataaca ggatcaaaaat 1320  
 agccccataaa aacgctaccg ttaa 1344

&lt;210&gt; 33

&lt;211&gt; 1911

&lt;212&gt; DNA

&lt;213&gt; Escherichia coli

&lt;400&gt; 33

gtgacggacg attttgcacc agacgggtcaag ctggcgaaag cgataccagg cttaagccg 60  
 cgagaaccac agcgacagat ggccgttagcc gtcaccagg cgatagaaaa aggccagccg 120  
 ctgggtgtgg aagcaggaac cggtaacgggg aaaacctacg cttacctggc tcctgcgtc 180  
 cgggcggaaa agaaaagtcat tatctcgacc ggctaaaag cgttgcagga tcagctctac 240  
 agccgcgatt tgccaaacagt ctcggaaaggca ttgaaatata cggcaacgt ggcgcgtgc 300  
 aaagggcgcct caaactacct ctgcctcgaa cgtctcgaa acgaggcgct ggcggggggc 360  
 gatctgcggc tacaatctt aagcgatgtt atcctgcgc gctctgtgc taatcaaaaca 420  
 gtcgatgggt atatcagcac ctgcgtcaggc gtggcgaaag attcacaggc gtggccgcgt 480  
 gtcaccaggca ccaacgacaa ctgtcttggc acgcgactgc cgatgtataa agattgttt 540  
 gtggcgaaag cacgtaaaaaa agcgatggac gcccgtatgg tgggtgtaaa ccatcatctc 600  
 ttctggcgat atatgggtt taaagagatgg ggatttgcg aactgatccc ggaagcggac 660  
 gtcatgtatct tcgacgaagc ccaccagcta cccgacatttgc ccagccagta ttgtgtcag 720  
 tcactctcca gtcgacaact gtcgacccgt gcaaaagaca tcaccatgc ctaccgcacc 780  
 gaattaaaaag acacccagca gttacaaaag tgcgtgtatc gtcttgcaca gagtgcgcag 840  
 gatttcgatc tgcaactcggt tgacccaggat ttcgcgtt acctgcgtga gctgttagct 900  
 aatccgaaaa ttccagggc attttacttgc ctcgatgaca ccctggaaact ttgttatgac 960  
 gtggcgaaac tgcgtactggg gctgtccgc ttgtgtatgc cggcatttgc ggcgcac 1020  
 ttgtatcgca cacggctgaa gcccgtaaaaa gagatcaatc agccggcata cagctactgg 1080  
 tacgaatgc cttccgcacc tttaactctgt gctctcacgc cgctcgcgt ggcggataaaa 1140  
 ttccaaagatgt taatggcgca aaaacccggt agctggatct tcacctcgcg aacgctgtcg 1200  
 gtgaacgcacg atctgcata tcacccatcg cggcttggca tcgaacaggc cgagtcgttg 1260  
 ctgttgcacc gcccatttgc ttacagccgc caggcgatc tctgtgtgc ggcgcatactg 1320  
 cccgcaaaacca accagccagg ttctgcgtc cagttgcgg caatgcgtcg accgatcatc 1380  
 gaagctaaaca acggtcgtt ttttatgtt tgtaacctgc acgcctatgt ggcgcgtctg 1440  
 gcccggcgt tccgcgtac catgacgctt cctgtattgt tgcaaggggg aaccagcaaa 1500  
 gggcaactgt tgcaccaatt tgcgtactggc ggtaatgcgc ttcttgcgc aaccagcaat 1560  
 ttctggcgaaag ggggtggacgt gctggcgat acattgtcat tggtaattat cgacaaattg 1620  
 ccgttaccc tccgcgtatgc tccactgtt aaagcgcgc tggaaagatgg tcgtttgcgc 1680  
 ggtggcgacc cgttcgatgc agtgcacta ccagatgcgc tcattactct caaacagggg 1740  
 gttagggcgac tgattcgcga cggcgcgtatc cgtggcgatc tggcgatgg tgacaatcg 1800

11

ctgggtatgc gtccttacgg cgcgacgtt ctcgcccagtc tgccgcggc gccacgcacc 1860  
cttgacattq cccatgtggat tcgtttctt qcqataccat cctccaggta a 1911

<210> 34  
<211> 414  
<212> DNA  
<213> Escherichia coli

```
<400> 34
atggctaata aacccgttgcg agaagaactg aaaaaaaaaatt tgtccgagat gcagttttac 60
gtgacgcaga atcatgggac agaaccgcga ttacgggtc gtttactgca taacaaggct 120
gacggcgat atcaactgttt gatctgcgtat gccccgcgtt ttcattttcca aaccaggat 180
gattcccgct gtggctggcc cagttttctac gaacccgtaa gtgaagaatc cattcggtt 240
atcaaaaggat tgcacatggc aatgcaggcgc atagaaaattt gttcggtta ctgtgtatggc 300
catctggggc atgtttttcccg cgcacggcccg cagccaaacgg gcaacgtta ttgtgttaaac 360
tctgcctt tacqctttac cgatggcgaa aacggcgaag aatcaacgg ttga 414
```

<210> 35  
<211> 1968  
<212> DNA  
<213> Escherichia coli

<400>	35					
atgaccattt	agaaaatttt	caccccgca	gacgacgcgt	tttatgcgg	gatcacccac	60
gcggcggggc	cgcaggcg	tctgcgcgt	accccgcaga	tgctgtatg	atctccac	120
ggcaacctgt	tcggcatgac	gcagaacg	ggatggct	gggacgcca	caagctcacc	180
ggcaaagagg	tgctgattat	cggaactcag	ggcggatcc	gcccggaga	ccggacgccc	240
atcgcgctgg	gttacacac	cgggatattgg	gagatcgca	tgcagatgc	ggcggcgcc	300
aaggagatca	ccccaaatgg	cgggatccc	ttcgccgcct	tcgtcagc	tccgtgcac	360
gggcgctcgc	agggacgca	cggtatgtt	gattccctgc	cgtaccgca	cgacgccc	420
atcggtttt	ggccgcgtat	cgctccctg	ccgacggc	ggggcgtat	ccgcgttagc	480
acctgcata	aagggtgtcc	cgccaccatg	attgcgtt	ccgcgtatgc	cgacctgccc	540
actattctgg	tgccggcg	ggcgacgt	ccggcacc	tcgggaaaga	cgccggcaag	600
gtgcagacca	tcggcgccg	tttcgcac	cacgaactt	ccctgcagga	geccgccc	660
ctgggctgtc	gcccgcgc	ctcgccgg	ggcggtgtc	agttctc	cacggcggc	720
acctcgca	tggtcgcg	ggcgctgg	ctggcgctc	cgcactcc	gctggcgcc	780
tccggcagg	cggtgtgg	ggagatcg	cgccagtc	cgcgcg	cagcgagtc	840
gatagccgc	gcataccac	gcccggat	ctctccgata	aaggccatgc	aaacgcgt	900
gtgatccac	cggtgttgc	cggttccacc	aatttactc	tgcacattc	gcccattc	960
cacgcccgg	gtgtcacat	ccccggat	gagacttgc	cgcgcata	cgttaaagt	1020
cccgctctt	tgagctgt	ggccaaac	ccggactatc	acccgacc	gcccgccttc	1080
ctcgccggc	cggtgcgc	gggtatgtc	cacctgcgc	acctcgcc	gctgcata	1140
gacgcata	ccgtgacc	ccagacgg	ggcgagaacc	ttgaatgt	gcaggcg	1200
gagcgcgg	cgcgcttc	ccagtgc	cgcgagc	acggcg	taga	1260
gtgatctgc	cgccggag	ggcaaaag	aaagggtgt	cctcgac	ctgttcc	1320
acggcaaca	tcgctccg	aggttgcgt	atcaaggc	cgcgatc	cccgctcg	1380
gtggcgaa	atggcgat	ccaccac	ggccgggt	gggtt	tctcgaaag	1440
caggcgatca	aggcgatc	gcccggag	attgtgc	gcgat	atcat	1500
ggcggcggc	cgtccggc	ggcatgg	gagacttac	agtcac	cgcgtaa	1560
catactctgt	ggggcaag	gggtgcgt	atcaccat	cgcgtt	cggcggtgc	1620
acggcgcc	gttctcgcc	cgtgtcg	gaggcg	cggcg	gattggca	1680
ctcgccgata	acgacatcat	cgagatgc	gtgatgtc	tgacgtt	ac	1740
aacttcatc	gcaccccg	caacccgc	acggcgg	gggcgc	cgagctgg	1800
cgccggcaga	cgcaccccg	cctgcac	cacgactt	tgccggac	cacccgg	1860
tggggccac	tgcgtcg	gagcggc	acctgg	gctgtat	ttat	1920
aaaattatcg	aggtat	cccggtaaa	aaagcgctc	gaattaa		1968

<210> 36  
<211> 717  
<212> DNA  
<213> Escherichia coli

<400> 36  
gtgggacgt aatggccaa tatttgtct aaaaaaacgg ctaaagacgg tgcaacgtct 60  
aaaatttgc caaaattcgg ttagaaatc tatgtctgtg ctaaacaagg tgaacccgat 120  
ccagaattaa acacatctt aaaattcgtt attgaacgt caaagcaggc acaagttcca 180  
aagcacgtt ttgataaagc aattgataaa gccaaaggcg gcggagatga aacggtcgtg 240  
cagggacgtt atgaaggctt tggtcctaatt ggctcaatga ttatcgccga gacattgact 300  
tcaaataatgtt accgtacgt tgctaacgtt cgccacaattt tcaataaaaa aggccgcaat 360  
atcgaggccg cagggtctgt cagctatatg tttgacaata cgggtgtat tttatattaa 420  
gggacagacc ctgaccatat tttgaaatt ttacttgaag ctgaagttga ttttcgtat 480  
gtgactgaag aagaaggtaa cattgttatt tataactgaac ctactgacct tcataaagga 540  
atcgccgctc taaaagcgc tggaatcact gagttctaa caacagaatt agaaatgatt 600  
gctcaatctg aagttagct ttcccagaa gatttagaaa tctttgaagg gcttggat 660  
gcccttgaag atgacgacga tgtaaaaaa gtttatcata acgtcgcaaa tctctaa 717

<210> 37  
<211> 258  
<212> DNA  
<213> Escherichia coli

<400> 37  
atggcgccgc cactgtcgcc tggctcgccg gtcctgatag ccctcattcg ggtctatcaa 60  
cgcctgatta gtccgctact cggccgcatt tgctgtttca ctccaaacctg ttcaagctac 120  
gaaatttgggg cattgcgcag gtttggatgt ataaaaggca gttgggtgac ggtgaaacgc 180  
gtattaaat gccaccctt acaccctgtt ggtgacgatc ccgtcccgcc cggaccattt 240  
gataccagag aacactaa 258

<210> 38  
<211> 1023  
<212> DNA  
<213> Escherichia coli

<400> 38  
atggaaaaag tgttattgtt aatcttgta ttgctgggtg tactgggtat cggcgctgg 60  
gtggcgctt ggaaggttcg ccatcttgc gacagcaat tgcttatcaa agaagagacg 120  
atatttaccc tgaaggccagg gaccggacgt ctggcgctcg gtgaacagct ttatgcccgt 180  
aagatcatca atcgccacg gttttttcaaa tggctgtgc gtatcgaaacc ggatctttct 240  
cactttaaaag ccggactta ccgttttaca ccgcagatga ccgtgcgcga gatgtgaaa 300  
ttgtggaaa gcggtttaaaa agcacagttc cctctgcgc tggtagaagg gatgcgtctg 360  
agcgattacc tcaagcaatt gctgtggggcc ccgtatataca agcatacgt gaggcatgtat 420  
aagtacgcca ccgtacgcga ggcacttggaa ctggaaaaacc cggagttggat tggatgg 480  
ttctggccag acacctggat gtataccgccc aataccaccc atgtcgctt actcaagcga 540  
gcccacaaga aaatggtcaa agccgtcgat agccctggg aaggccgtgc ggacggcttg 600  
ccttataaaag ataaaaaccat gttggtgacg atggcatcaa ttatcgaaaa agaaaccgccc 660  
gttgcctgtt aacgcgatataa ggttgcctca gtatttatca accgtttacg cattggatgt 720  
ccgcctgcaga ccgaccgcac cgtgatttac gggatggggag agcggttataa tggcaaaactt 780  
tctcgctgcag acctgaaac gccgacagcg tataacacctt ataccattac cgtctgcgc 840  
ccaggtgcga taggcacgc gggggccggat tcgctgaagg ctgctgcga tccggcaaaa 900  
acgcccgtatc tctatattgtt gggcgatgtt aaaggtggtc acacgtttaa taccaatctt 960  
gccagtcata acaagtctgtt gcaggattat ctgaaagtgc ttaaggaaaa aatgcgcag 1020  
taa 1023

<210> 39  
<211> 873  
<212> DNA  
<213> Escherichia coli

```

<400> 39
atgctgcggc actcatcagt ccgtttaaat aaatacatca gcgaaagcg  aatttgctca 60
cgccgcgaag cggatcgcta tatcgagcaa ggcaatgtgt tccttaatgg caagcgaccc 120
accattggcg atcagggtgaa acccggcgac gttgtgaaaag taaaacggtca gttgattgaa 180
ccctcgaaaaag ccgaagattt ggtactttatc gccctgaaca agcccggtgg tattgttaagc 240
accaccgaag atggcgagcg cgataacatt gtcgatttcg ttaaccacag caaacgcgtg 300
ttcccgattt gcccgccttgg taagacttcc caggggttga ttttccctcac caatcacggc 360
gatctggtga ataagatctt gctgtcttgg aatgatcatg agaaagagta tctgtgtgacg 420
gtcgataaac cgattaccga ggagtttattt cgccggatgtg gtgggggggtt gccaatccccc 480
gggacagtga cccaaaaagtgc caaagttaaaa aaagaagcgcc cggttgtctt ccgcattacc 540
ctgtgtcagg ggcttgcaccc tcagatcccg cgcatgtgcg agcatttccgg ctatgaagtgc 600
aaaaagctgg aacgcacgcg catcatgaaac gtttagcttaa ggcgcatttcc gctggggggaa 660
tggcgcgtttaaaccgcga tgtagttaatc gaccccttta agtcatttgc aaatttcctct 720
tccgaggtaa aacctaaagc gaaggccaaa ccgaaaaacag cgggcatttca acgttccagtc 780
gttaagatgg aaaaaacggc ggaaaaaaggc gggtcgccccgg cgtccaaacgg taagcgtttt 840
acctcgccgg ggcgtaaaaaa gaagggggcgcc tga 873

```

<210> 40  
<211> 159  
<212> PRT  
<213> Escherichia coli

<400> 40  
 Met Arg Ile Gly His Gly Phe Asp Val His Ala Phe Gly Gly Glu Gly  
 1 5 10 15  
 Pro Ile Ile Ile Gly Gly Val Arg Ile Pro Tyr Glu Lys Gly Leu Leu  
 20 25 30  
 Ala His Ser Asp Gly Asp Val Ala Leu His Ala Leu Thr Asp Ala Leu  
 35 40 45  
 Leu Gly Ala Ala Ala Leu Gly Asp Ile Gly Lys Leu Phe Pro Asp Thr  
 50 55 60  
 Asp Pro Ala Phe Lys Gly Ala Asp Ser Arg Glu Leu Leu Arg Glu Ala  
 65 70 75 80  
 Trp Arg Arg Ile Gln Ala Lys Gly Tyr Thr Leu Gly Asn Val Asp Val  
 85 90 95  
 Thr Ile Ile Ala Gln Ala Pro Lys Met Leu Pro His Ile Pro Gln Met  
 100 105 110  
 Arg Val Phe Ile Ala Glu Asp Leu Gly Cys His Met Asp Asp Val Asn  
 115 120 125  
 Val Lys Ala Thr Thr Thr Glu Lys Leu Gly Phe Thr Gly Arg Gly Glu  
 130 135 140  
 Gly Ile Ala Cys Glu Ala Val Ala Leu Leu Ile Lys Ala Thr Lys  
 145 150 155

<210> 41  
<211> 158  
<212> PRT  
<213> Haemophilus influenzae

<400> 41  
Met Ile Arg Ile Gly His Gly Phe Asp Val His Ala Phe Gly Glu Asp  
1 5 10 15  
Arg Pro Leu Ile Ile Gly Gly Val Glu Val Pro Tyr His Thr Gly Phe  
20 25 30  
Ile Ala His Ser Asp Gly Asp Val Ala Leu His Ala Leu Thr Asp Ala  
35 40 45  
Ile Leu Gly Ala Ala Ala Leu Gly Asp Ile Gly Lys Leu Phe Pro Asp  
50 55 60  
Thr Asp Met Gln Tyr Lys Asn Ala Asp Ser Arg Gly Leu Leu Arg Glu  
65 70 75 80  
Ala Phe Arg Gln Val Gln Glu Lys Gly Tyr Lys Ile Gly Asn Val Asp  
85 90 95  
Ile Thr Ile Ile Ala Gln Ala Pro Lys Met Arg Pro His Ile Asp Ala  
100 105 110  
Met Arg Ala Lys Ile Ala Glu Asp Leu Gln Cys Asp Ile Glu Gln Val  
115 120 125  
Asn Val Lys Ala Thr Thr Glu Lys Leu Gly Phe Thr Gly Arg Gln  
130 135 140  
Glu Gly Ile Ala Cys Glu Ala Val Ala Leu Ile Arg Gln  
145 150 155

<210> 42  
<211> 158  
<212> PRT  
<213> Bacillus subtilis

<400> 42  
Met Phe Arg Ile Gly Gln Gly Phe Asp Val His Gln Leu Val Glu Gly  
1 5 10 15  
Arg Pro Leu Ile Ile Gly Gly Ile Glu Ile Pro Tyr Glu Lys Gly Leu  
20 25 30  
Leu Gly His Ser Asp Ala Asp Val Leu Leu His Thr Val Ala Asp Ala  
35 40 45  
Cys Leu Gly Ala Val Gly Glu Gly Asp Ile Gly Lys His Phe Pro Asp  
50 55 60  
Thr Asp Pro Glu Phe Lys Asp Ala Asp Ser Phe Lys Leu Leu Gln His  
65 70 75 80

Val Trp Gly Ile Val Lys Gln Lys Gly Tyr Val Leu Gly Asn Ile Asp  
85 90 95

Cys Thr Ile Ile Ala Gln Lys Pro Lys Met Leu Pro Tyr Ile Glu Asp  
100 105 110

Met Arg Lys Arg Ile Ala Glu Gly Leu Glu Ala Asp Val Ser Gln Val  
115 120 125

Asn	Val	Lys	Ala	Thr	Thr	Thr	Glu	Lys	Leu	Gly	Phe	Thr	Gly	Arg	Ala
130							135							140	

Glu Gly Ile Ala Ala Gln Ala Thr Val Leu Ile Gln Lys Gly  
145 150 155

<210> 43  
<211> 161  
<212> PRT  
<213> *Synechocystis* sp.

<400> 43  
Met Thr Ala Leu Arg Ile Gly Asn Gly Tyr Asp Ile His Arg Leu Val  
1 5 10 15

Gly Asp Arg Pro Leu Ile Leu Gly Gly Val Thr Ile Ala His His Leu  
20 25 30

Gly Leu Asp Gly His Ser Asp Ala Asp Val Leu Thr His Ala Leu Met  
35 40 45

Asp Ala Leu Leu Gly Ala Leu Ser Leu Gly Asp Ile Gly His Tyr Phe  
50 55 60

Pro Pro Ser Asp Ala Arg Trp Gln Gly Ala Asp Ser Leu Lys Leu Leu  
 65                    70                    75                    80

Ala Gln Val His Gln Leu Ile Leu Glu Arg Gly Trp Arg Ile Asn Asn  
85 90 95

Leu Asp Asn Val Ile Val Ala Glu Gln Pro Lys Leu Lys Pro His Ile  
100 105 110

Gln Ala Met Lys Glu Asn Leu Ala Lys Val Leu Thr Ile Asp Pro Asp  
115 120 125

Leu Ile Gly Ile Lys Ala Thr Thr Asn Glu Arg Leu Gly Pro Thr Gly  
130 135 140

Arg Glu Glu Gly Ile Ala Ala Tyr Ser Val Ala Leu Leu Ile Lys Glu  
 145                    150                    155                    160

Gly

<210> 44  
<211> 399

16

&lt;212&gt; PRT

&lt;213&gt; Treponema pallidum

&lt;400&gt; 44

Met Arg Arg Gly Gly Ala Cys Val Gln Lys Lys Glu Tyr Leu Pro Leu  
1 5 10 15

Thr Ser Arg Gln Pro Gly Val Cys Leu Leu Ser Glu Ile Leu Val Arg  
20 25 30

Ala Leu Glu Ala Arg Ser Phe Phe Leu Val Val Val Thr Val Pro Ala  
35 40 45

Gly Glu Val Ala Tyr Ala Glu Ser Gln Val Ala Cys Asp Ser Arg Leu  
50 55 60

Ser Ala Phe Pro Ser Arg Thr Arg Pro Val Ile Leu Tyr Val Pro Gly  
65 70 75 80

Ala His Thr Arg Ser Ala Ser Val Arg Ala Gly Leu Asp Ala Met Ala  
85 90 95

Thr His Ala Pro Asp Val Val Leu Val His Asp Gly Ala Arg Pro Phe  
100 105 110

Val Ser Val Ala Leu Ile His Ser Val Leu Glu Ala Thr Cys Arg Tyr  
115 120 125

Gly Ala Ala Val Pro Val Ile Glu Ala Thr Asp Thr Pro Lys Gly Val  
130 135 140

Ala Ala Asp Gly Ser Ile Glu Thr His Leu Ile Arg Ser Arg Val Arg  
145 150 155 160

Leu Ala Gln Thr Pro Gln Gly Phe Cys Tyr Ala Ser Leu Cys Ala Ala  
165 170 175

His His Arg Ala Ala Thr Asp Gly Glu Gln Tyr Thr Asp Asp Ser Glu  
180 185 190

Leu Tyr Ala Arg Tyr Gly Gly Thr Val His Val Cys Ala Gly Glu Arg  
195 200 205

Ser Asn Val Lys Ile Thr Tyr Pro Glu Asp Leu Glu Gln Arg Ala Ser  
210 215 220

Glu Pro Ala Leu Thr Arg Gly Ile Ser Val Leu Pro Cys Thr Glu Glu  
225 230 235 240

Gly Ala Leu Arg Val Gly Leu Gly Thr Asp Met His Ala Leu Cys Ala  
245 250 255

Gly Arg Pro Leu Ile Leu Ala Gly Ile His Ile Pro Ser Lys Lys Gly  
260 265 270

Ala Gln Gly His Ser Asp Ala Asp Val Leu Ala His Ala Ser Ile Asp  
275 280 285

Ala Leu Leu Gly Ala Ala Gly Leu Gly Asp Ile Gly Thr Phe Pro

17

290	295	300
Ser Cys Asp Gly Arg Trp Lys Asp Ala His Ser Cys Ala Leu Leu Arg		
305	310	315
His Thr Trp Gln Leu Val Arg Ala Ala Cys Trp Arg Leu Val Asn Leu		
325	330	335
Asp Ala Val Val Cys Leu Glu Gln Pro Ala Leu His Pro Phe Arg Glu		
340	345	350
Ala Met Arg Ala Ser Leu Ala Gln Ala Leu Asp Thr His Val Thr Arg		
355	360	365
Val Phe Val Lys Ala Lys Thr Ala Glu Arg Leu Gly Pro Val Gly Ser		
370	375	380
Gly Ala Ala Val Thr Ala Gln Val Val Val Leu Leu Lys Lys Ile		
385	390	395
<210> 45		
<211> 406		
<212> PRT		
<213> Helicobacter pylori		
<400> 45		
Met Ser Leu Ile Arg Val Asn Gly Glu Ala Phe Lys Leu Ser Leu Glu		
1	5	10
Ser Leu Glu Glu Asp Pro Phe Glu Thr Lys Glu Thr Leu Glu Thr Leu		
20	25	30
Ile Lys Gln Thr Ser Val Val Leu Leu Ala Ala Gly Glu Ser Arg Arg		
35	40	45
Phe Ser Gln Thr Ile Lys Lys Gln Trp Leu Arg Ser Asn His Thr Pro		
50	55	60
Leu Trp Leu Ser Val Tyr Glu Ser Phe Lys Glu Ala Leu Asp Phe Lys		
65	70	75
80		
Glu Ile Ile Leu Val Val Ser Glu Leu Asp Tyr Ile Tyr Ile Lys Arg		
85	90	95
His Tyr Pro Glu Ile Lys Leu Val Lys Gly Gly Ala Ser Arg Gln Glu		
100	105	110
Ser Val Arg Asn Ala Leu Lys Ile Ile Asp Ser Ala Tyr Thr Leu Thr		
115	120	125
Ser Asp Val Ala Arg Gly Leu Ala Asn Ile Glu Ala Leu Lys Asn Leu		
130	135	140
Phe Leu Thr Leu Gln Gln Thr Ser His Tyr Cys Ile Ala Pro Tyr Leu		
145	150	155
160		
Pro Cys Tyr Asp Thr Ala Ile Tyr Tyr Asn Glu Ala Leu Asp Arg Glu		
165	170	175

Ala Ile Lys Leu Ile Gln Thr Pro Gln Leu Ser His Thr Lys Ala Leu  
180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400  
Gln Ser Ala Leu Asn Gln Gly Asp Phe Lys Asp Glu Ser Ser Ala Ile  
Leu Gln Ala Phe Pro Asp Arg Val Ser Tyr Ile Glu Gly Ser Lys Asp  
Leu His Lys Leu Thr Thr Ser Gly Asp Leu Lys His Phe Thr Leu Phe  
Phe Asn Pro Ala Lys Asp Thr Phe Ile Gly Met Gly Phe Asp Thr His  
Ala Phe Ile Lys Asp Pro Met Val Leu Gly Gly Val Val Leu Asp  
Cys Glu Phe Gly Leu Lys Ala His Ser Asp Gly Asp Ala Leu Leu His  
Ala Val Ile Asp Ala Ile Leu Gly Ala Ile Lys Gly Gly Asp Ile Gly  
Glu Trp Phe Pro Asp Asn Asp Pro Lys Tyr Lys Asn Ala Ser Ser Lys  
Glu Leu Leu Lys Ile Val Leu Asp Phe Ser Gln Ser Ile Gly Phe Glu  
Leu Phe Glu Met Gly Ala Thr Ile Phe Ser Glu Ile Pro Lys Ile Thr  
Pro Tyr Lys Pro Ala Ile Leu Glu Asn Leu Ser Gln Leu Leu Gly Leu  
Glu Lys Ser Gln Ile Ser Leu Lys Ala Thr Thr Met Glu Lys Met Gly  
Phe Ile Gly Lys Gln Glu Gly Leu Leu Val Gln Ala His Val Ser Met  
Arg Tyr Lys Gln Lys Leu